~~TRANSACYLASES OF THE PACLITAXEL BIOSYNTHETIC PATHWAY~~ Nucleic Acid Molecules Encoding 10-Deacetylbaccatin III-O-Acetyl Transferase and Related Products

FIELD OF THE INVENTION

The invention relates to transacylase enzymes and methods of using such enzymes to produce Taxol™ and related taxoids.

CROSS REFERENCE TO RELATED CASES

This application is a continuation in part of co-pending U.S. Application No. 09/411,145, filed September 30, 1999, which is incorporated herein by reference.

ACKNOWLEDGMENT OF GOVERNMENT SUPPORT

This invention was made with government support under National Cancer Institute Grant No. CA-55254. The government has certain rights in this invention.

INTRODUCTION

The complex diterpenoid Taxol™ (paclitaxel) (Wani et al., *J. Am. Chem. Soc.* **93**:2325-2327, 1971) is a potent antimitotic agent with excellent activity against a wide range of cancers, including ovarian and breast cancer (Arbuck and Blaylock, *Taxol™: Science and Applications*, CRC Press, Boca Raton, 397-415, 1995; Holmes et al., *ACS Symposium Series* **583**:31-57, 1995). Taxol™ was originally isolated from the bark of the Pacific yew (*Taxus brevifolia*). For a number of years, Taxol™ was obtained exclusively from yew bark, but low yields of this compound from the natural source coupled to the destructive nature of the harvest, prompted new methods of Taxol™ production to be developed. Taxol™ is currently produced primarily by semisynthesis from advanced taxane metabolites (Holton et al., *Taxol™: Science and Applications*, CRC Press, Boca Raton, 97-121, 1995) that are present in the needles (a renewable resource) of various *Taxus* species. However, because of the increasing demand for this drug (both for use earlier in the course of cancer intervention and for new therapeutic applications) (Goldspiel, *Pharmacotherapy* **17**:110S-125S, 1997), availability and cost remain important issues. Total chemical synthesis of Taxol™ is not economically feasible. Hence, biological production of the drug and its immediate precursors will remain the method of choice for the foreseeable future. Such biological production may rely upon either intact *Taxus* plants, *Taxus* cell cultures (Ketchum et al., *Biotechnol. Bioeng.* **62**:97-105, 1999), or, potentially, microbial systems (Stierle et al., *J. Nat. Prod.* **58**:1315-1324,

1995). In all cases, improving the biological production yields of Taxol depends upon a detailed understanding of the biosynthetic pathway, the enzymes catalyzing the sequence of reactions, especially the rate-limiting steps, and the genes encoding these proteins. Isolation of genes encoding enzymes involved in the pathway is a particularly important goal, since overexpression of these genes in a producing organism can be expected to markedly improve yields of the drug.

The Taxol™ biosynthetic pathway is considered to involve more than 12 distinct steps (Floss and Mocek, *Taxol: Science and Applications*, CRC Press, Boca Raton, 191-208, 1995; and Croteau et al., *Curr. Top. Plant Physiol.* **15**:94-104, 1996), however, very few of the enzymatic reactions and intermediates of this complex pathway have been defined. The first committed enzyme of the Taxol™ pathway is taxadiene synthase (Koepp et al., *J. Biol. Chem.* **270**:8686-8690, 1995) that cyclizes the common precursor geranylgeranyl diphosphate (Hefner et al., *Arch. Biochem. Biophys.* **360**:62-74, 1998) to taxadiene (**Figure 1**). The cyclized intermediate subsequently undergoes modification involving at least eight oxygenation steps, a dehydrogenation, an epoxide rearrangement to an oxetane, and several acylations (Floss and Mocek, *Taxol™: Science and Applications*, CRC Press, Boca Raton, 191-208, 1995; Croteau et al., *Curr. Top. Plant Physiol.* **15**:94-104, 1996). Taxadiene synthase has been isolated from *T. brevifolia* and characterized (Hezari et al., *Arch. Biochem. Biophys.* **322**:437-444, 1995), the mechanism of action defined (Lin et al., *Biochemistry* **35**:2968-2977, 1996), and the corresponding cDNA clone isolated and expressed (Wildung and Croteau, *J. Biol. Chem.* **271**:9201-9204, 1996).

Sub C3 The second specific step of Taxol™ biosynthesis is an oxygenation reaction catalyzed by taxadiene-5∀-hydroxylase (**Figure 1**). The enzyme, characterized as a cytochrome P450, has been demonstrated in *Taxus* microsome preparations to catalyze the stereospecific hydroxylation of taxa-4(5),11(12)-diene, with double bond rearrangement, to taxa-4(20),11(12)-dien-5∀-ol (Hefner et al., *Chem. Biol.* **3**:479-489, 1996).

Sub C4 The third specific step of Taxol™ biosynthesis appears to be the acetylation of taxa-4(20),11(12)-dien-5∀-ol to taxa-4(20),11(12)-dien-5∀-yl acetate by an acetyl CoA-dependent transacetylase (Walker et al., *Arch. Biochem. Biophys.* **364**:273-279, 1999), since the resulting acetate ester is then further efficiently oxygenated to a series of advanced polyhydroxylated Taxol™ metabolites in microsomal preparations that have

been optimized for cytochrome P450 reactions (**Figure 1**). The enzyme has been isolated from induced yew cell cultures (*Taxus canadensis* and *Taxus cuspidata*), and the operationally soluble enzyme was partially purified by a combination of anion exchange, hydrophobic interaction, and affinity chromatography on immobilized coenzyme A resin. This acetyl transacylase has a pI and pH optimum of 4.7 and 9.0, respectively, and a molecular weight of about 50,000 as determined by gel-permeation chromatography. The enzyme shows high selectivity and high affinity for both cosubstrates with K_m values of 4.2 μM and 5.5 μM for taxadienol and acetyl CoA, respectively. The enzyme does not acetylate the more advanced Taxol™ precursors, 10-deacetylbaccatin III or baccatin III. This acetyl transacylase is insensitive to monovalent and divalent metal ions, is only weakly inhibited by thiol-directed reagents and Co-enzyme A, and in general displays properties similar to those of other *O*-acetyl transacylases. This acetyl CoA:taxadien-5∀-ol *O*-acetyl transacylase from *Taxus* (Walker et al., *Arch. Biochem. Biophys.* **364**:273-279, 1999) appears to be substantially different in size, substrate selectivity, and kinetics from an acetyl CoA:10-hydroxytaxane *O*-acetyl transacylase recently isolated and described from *Taxus chinensis* (Menhard and Zenk, *Phytochemistry* **50**:763-774, 1999). Acquisition of the gene encoding the acetyl CoA:taxa-4(20),11(12)-dien-5∀-ol *O*-acetyl transacylase that catalyzes the first acylation step of Taxol™ biosynthesis and genes encoding other acyl transfer steps would represent an important advance in efforts to increase Taxol™ yields by genetic engineering and *in vitro* synthesis.

SUMMARY OF THE INVENTION

The invention stems from the discovery of twelve amplicons (regions of DNA amplified by a pair of primers using the polymerase chain reaction (PCR)). These amplicons can be used to identify transacylases, for example, the transacylases shown in SEQ ID NOs: 26, 28, 45, 50, 52, 54, 56, and 58 that are encoded by the nucleic acid sequences shown in SEQ ID NOs: 25, 27, 44, 49, 51, 53, 55, and 57. These sequences are isolated from the *Taxus* genus, and the respective transacylases are useful for the synthetic production of Taxol™ and related taxoids, as well as intermediates within the Taxol™ biosynthetic pathway. The sequences can be also used for the creation of transgenic organisms that either produce the transacylases for subsequent *in vitro* use, or produce the transacylases *in vivo* so as to alter the level of Taxol™ and taxoid production within the transgenic organism.

Another aspect of the invention provides the nucleic acid sequences shown in SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17; 19, 21, and 23 and the corresponding amino acid sequences shown in SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24, respectively, as well as fragments of the nucleic acid and the amino acid sequences. These sequences are useful for isolating the nucleic acid and amino acid sequences corresponding to full-length transacylases. These amino acid sequences and nucleic acid sequences are also useful for creating specific binding agents that recognize the corresponding transacylases.

Accordingly, another aspect of the invention provides for the identification of transacylases and fragments of transacylases that have amino acid and nucleic acid sequences that vary from the disclosed sequences. For example, the invention provides transacylase amino acid sequences that vary by one or more conservative amino acid substitutions, or that share at least 50% sequence identity with the amino acid sequences provided while maintaining transacylase activity.

The nucleic acid sequences encoding the transacylases and fragments of the transacylases can be cloned, using standard molecular biology techniques, into vectors. These vectors can then be used to transform host cells. Thus, a host cell can be modified to express either increased levels of transacylase or decreased levels of transacylase.

Another aspect of the invention provides methods for isolating nucleic acid sequences encoding full-length transacylases. The methods involve hybridizing at least ten contiguous nucleotides of any of the nucleic acid sequences shown in SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 44, 49, 51, 53, 55, and 57 to a second nucleic acid sequence, wherein the second nucleic acid sequence encodes a transacylase. This method can be practiced in the context of, for example, Northern blots, Southern blots, and the polymerase chain reaction (PCR). Hence, the invention also provides the transacylases identified by this method.

Yet another aspect of the invention involves methods of adding at least one acyl group to at least one taxoid. These methods can be practiced *in vivo* or *in vitro*, and can be used to add acyl groups to various intermediates in the Taxol™ biosynthetic pathway, and to add acyl groups to related taxoids that are not necessarily in a Taxol™ biosynthetic pathway.

SEQUENCE LISTINGS

The nucleic and amino acid sequences listed in the accompanying sequence listing are shown using standard letter abbreviations for nucleotide bases, and three-letter code for amino acids. Only one strand of each nucleic acid sequence is shown, but the complementary strand is understood to be included by any reference to the displayed strand.

SEQ ID NO: 1 is the nucleotide sequence of Probe 1.

SEQ ID NO: 2 is the deduced amino acid sequence of Probe 1.

SEQ ID NO: 3 is the nucleotide sequence of Probe 2.

SEQ ID NO: 4 is the deduced amino acid sequence of Probe 2.

SEQ ID NO: 5 is the nucleotide sequence of Probe 3.

SEQ ID NO: 6 is the deduced amino acid sequence of Probe 3.

SEQ ID NO: 7 is the nucleotide sequence of Probe 4.

SEQ ID NO: 8 is the deduced amino acid sequence of Probe 4.

SEQ ID NO: 9 is the nucleotide sequence of Probe 5.

SEQ ID NO: 10 is the deduced amino acid sequence of Probe 5.

SEQ ID NO: 11 is the nucleotide sequence of Probe 6.

SEQ ID NO: 12 is the deduced amino acid sequence of Probe 6.

SEQ ID NO: 13 is the nucleotide sequence of Probe 7.

SEQ ID NO: 14 is the deduced amino acid sequence of Probe 7.

SEQ ID NO: 15 is the nucleotide sequence of Probe 8.

SEQ ID NO: 16 is the deduced amino acid sequence of Probe 8.

SEQ ID NO: 17 is the nucleotide sequence of Probe 9.

SEQ ID NO: 18 is the deduced amino acid sequence of Probe 9.

SEQ ID NO: 19 is the nucleotide sequence of Probe 10.

SEQ ID NO: 20 is the deduced amino acid sequence of Probe 10.

SEQ ID NO: 21 is the nucleotide sequence of Probe 11.

SEQ ID NO: 22 is the deduced amino acid sequence of Probe 11.

SEQ ID NO: 23 is the nucleotide sequence of Probe 12.

SEQ ID NO: 24 is the deduced amino acid sequence of Probe 12.

SEQ ID NO: 25 is the nucleotide sequence of the full-length acyltransacylase clone TAX2.

SEQ ID NO: 26 is the deduced amino acid sequence of the full-length acyltransacylase clone TAX2.

SEQ ID NO: 27 is the nucleotide sequence of the full-length acyltransacylase clone TAX1.

SEQ ID NO: 28 is the deduced amino acid sequence of the full-length acyltransacylase clone TAX1.

SEQ ID NO: 29 is the amino acid sequence of a transacylase peptide fragment.

SEQ ID NO: 30 is the amino acid sequence of a transacylase peptide fragment.

SEQ ID NO: 31 is the amino acid sequence of a transacylase peptide fragment.

SEQ ID NO: 32 is the amino acid sequence of a transacylase peptide fragment.

SEQ ID NO: 33 is the amino acid sequence of a transacylase peptide fragment.

SEQ ID NO: 34 is the AT-FOR1 PCR primer.

SEQ ID NO: 35 is the AT-FOR2 PCR primer.

SEQ ID NO: 36 is the AT-FOR3 PCR primer.

SEQ ID NO: 37 is the AT-FOR4 PCR primer.

SEQ ID NO: 38 is the AT-REV1 PCR primer.

SEQ ID NO: 39 is an amino acid sequence variant that allowed for the design of the AT-FOR3 PCR primer.

SEQ ID NO: 40 is an amino acid sequence variant that allowed for the design of the AT-FOR4 PCR primer.

SEQ ID NO: 41 is a consensus amino acid sequence that allowed for the design of the AT-REV1 PCR primer.

SEQ ID NO: 42 is a PCR primer, useful for identifying transacylases.

SEQ ID NO: 43 is a PCR primer, useful for identifying transacylases.

SEQ ID NO: 44 is the nucleotide sequence of the full-length acyltransacylase clone TAX6.

SEQ ID NO: 45 is the deduced amino acid sequence of the full-length acyltransacylase clone TAX6.

SEQ ID NO: 46 is a PCR primer, useful for identifying TAX6.

SEQ ID NO: 47 is a PCR primer, useful for identifying TAX6.

SEQ ID NO: 48 is a 6-amino acid motif commonly found in transacylases.

SEQ ID NO: 49 is the nucleotide sequence of the full-length acyltransacylase clone TAX5.

SEQ ID NO: 50 is the deduced amino acid sequence of the full-length acyltransacylase clone TAX5.

SEQ ID NO: 51 is the nucleotide sequence of the full-length acyltransacylase clone TAX7.

SEQ ID NO: 52 is the deduced amino acid sequence of the full-length acyltransacylase clone TAX7.

SEQ ID NO: 53 is the nucleotide sequence of the full-length acyltransacylase clone TAX10.

SEQ ID NO: 54 is the deduced amino acid sequence of the full-length acyltransacylase clone TAX10.

SEQ ID NO: 55 is the nucleotide sequence of the full-length acyltransacylase clone TAX12.

SEQ ID NO: 56 is the deduced amino acid sequence of the full-length acyltransacylase clone TAX12.

SEQ ID NO: 57 is the nucleotide sequence of the full-length acyltransacylase clone TAX13.

SEQ ID NO: 58 is the deduced amino acid sequence of the full-length acyltransacylase clone TAX13.

FIGURES

Figure 1: Enzymatic reactions of the Taxol™ pathway indicating cyclization of geranylgeranyl diphosphate to taxa-4(5),11(12)-diene, followed by hydroxylation/rearrangement and acetylation to taxa-4(20),11(12)-dien-5α-yl acetate. The acetate is further converted to 10-deacetylbaccatin III, baccatin III, and Taxol™. In the figure, "a" denotes taxadiene synthase; "b" denotes taxadiene-5α-hydroxylase; "c" denotes taxadien-5α-ol acetyl transacylase; and "d" denotes several subsequent steps.

Figure 2: Peptide sequences generated by endolysC and trypsin proteolysis of purified taxadienol acetyl transacylase.

Figures 3A-3D

~~**Figure 3:**~~ Panel **A** is an elution profile of the acetyl transacylase on Source HR 15Q (10 X 100 mm) preparative scale anion-exchange chromatography; Panel **B** is an elution profile on analytical scale Source HR 15Q (5 X 50 mm) column chromatography; and Panel **C** is an elution profile on the ceramic hydroxyapatite column. The solid line is the UV absorbance at 280 nm; the dotted line is the relative transacetylase activity (dpm); and the hatched line is the elution gradient (sodium chloride or sodium phosphate). Panel **D** is a photograph of a silver-stained 12% SDS-PAGE showing the purity of taxadien-5α-ol acetyl transacylase (50 kDa) after hydroxyapatite chromatography. A minor contaminant is present at ~35 kDa.

Figure 4 shows four forward (AT-FOR1, AT-FOR2, AT-FOR3, AT-FOR4) and one reverse (AT-REV1) degenerate primers that were used to amplify an induced *Taxus* cell library cDNA from which twelve hybridization probes were obtained. Inosine positions are indicated by "I". Each of the forward primers was paired with the reverse primer in separate PCR reactions. Primers AT-FOR1 (SEQ ID NO: 34) and AT-FOR2 (SEQ ID NO: 35) were designed from the tryptic fragment SEQ ID NO: 30; the remaining primers were derived by database searching based on SEQ ID NO: 30.

KK 1/13/03

Figures 5A-5G show

~~Figure 5 shows~~ data obtained from a coupled gas chromatographic-mass spectrometric (GC-MS) analysis of the biosynthetic taxadien-5α-yl acetate formed during the incubation of taxadien-5α-ol with soluble enzyme extracts from isopropyl β-D-thiogalactoside (IPTG)-induced *E. coli* JM109 cells transformed with full-length acyltransacylase clones TAX1 and TAX2. Panels **A** and **B** show the respective GC and MS profiles of authentic taxadien-5α-ol; panels **C** and **D** show the respective GC and MS profiles of authentic taxadien-5α-yl acetate; panel **E** shows the GC profile of taxadien-5α-ol (11.16 minutes), taxadien-5α-yl acetate (11.82 minutes), dehydrated taxadien-5α-ol ("TOH-H_2O" peak), and a contaminant, bis-(2-ethylhexyl)phthlate ("BEHP" peak, a plasticizer, CAS 117-81-7, extracted from buffer) after incubation of taxadien-5α-ol and acetyl coenzyme A with the soluble enzyme fraction derived from *E. coli* JM109 transformed with the full-length clone TAX1. Panel **F** shows the mass spectrum of biosynthetically formed taxadien-5α-yl acetate by the recombinant enzyme (11.82 minute peak in GC profile Panel E); panel **G** shows the GC profile of the products generated from taxadien-5α-ol and acetyl coenzyme A by incubation with the soluble enzyme fraction derived from *E. coli* JM109 cells transformed with the full-length clone TAX2 (note the absence of taxadien-5α-yl acetate indicating that this clone is inactive in the transacylase reaction).

Figures 6A-6N

KK 1/13/03

~~Figure 6:~~ Pileup of deduced amino acid sequences listed in Table 1, and of TAX1 and TAX2. Residues boxed in black (and gray) indicate the few regions of conservation. Forward arrow (left to right) shows conserved region from which degenerate forward PCR primers were designed. Reverse arrow (right to left) shows region from which the reverse PCR primer was designed (cf., **Figure 4**).

Figure 7: Dendrogram showing deduced peptide sequence relationships between *Taxus* transacylase sequences (Probes 1-12, TAX1, and TAX2) and closest relative sequences of defined and unknown function obtained from the GenBank database described in Table 1.

Figures 8A and 8B

KK 1/13/03

~~Figure 8:~~ Panel **A** shows the outline of the Taxol™ biosynthetic pathway. The cyclization of geranylgeranyl diphosphate to taxadiene by taxadiene synthase, and the hydroxylation to taxadien-5α-ol by taxadiene 5α-hydroxylase (a), the acetylation of

taxadien-5α-ol by taxa-4(20),11(12)-dien-5α-ol-*O*-acetyl transferase (b), the conversion of 10-deacetylbaccatin III to baccatin III by 10-deacetylbaccatin III-10-*O*-acetyl transferase (c), and the side chain attachment to baccatin III to form Taxol™ (d) are highlighted. The broken arrow indicates several as yet undefined steps. Panel **B** shows a postulated biosynthetic scheme for the formation of the oxetane, present in Taxol™ and related late-stage taxoids, in which the 4(20)-ene-5α-ol is converted to the 4(20)-ene-5α-yl acetate followed by epoxidation to the 4(20)-epoxy-5α-acetoxy group and then intramolecular rearrangement to the 4-acetoxy oxetane moiety.

KK 1/13/03

Figures 9A and 9B

~~Figure 9~~: Radio-HPLC (high-performance liquid chromatography) analysis of the biosynthetic product (Rt = 7.0 ± 0.1 minutes) generated from 10-deacetylbaccatin III and [2-^{3}H]acetyl CoA by the recombinant acetyl transferase. The top trace shows the UV profile and the bottom trace shows the coincident radioactivity profile, both of which coincide with the retention time of authentic baccatin III. For the enzyme preparation,

E. coli cells transformed with the pCWori+ vector harboring the putative DBAT gene were grown overnight at 37°C in 5 mL Luria-Bertani medium supplemented with ampicillin, and 1 mL of this inoculum was added to and grown in 100 mL Terrific Broth culture medium (6 g bacto-tryptone, Difco Laboratories, Spark, MD, 12 g yeast extract, EM Science, Cherryhill, NJ, and 2 mL gycerol in 500 mL water) supplemented with 1 mM IPTG, 1 mM thiamine HCl and 50 μg ampicillin/mL. After 24 hours, the bacteria were harvested by centrifugation, resuspended in 20 mL of assay buffer (25 mM Mopso, pH 7.4) and then disrupted by sonication at 0-4°C. The resulting homogenate was centrifuged at 15,000 *g* to remove debris, and a 1 mL aliquot of the supernatant was incubated with 10-deacetylbaccatin III (400 μM) and [2-^{3}H]acetyl coenzyme A (0.45 μCi, 400 μM) for 1 hour at 31°C. The reaction mixture was then extracted with ether and the solvent concentrated *in vacuo*. The crude product (pooled from five such assays) was purified by silica gel thin-layer chromatography (TLC; 70:30 ethyl acetate: hexane). The band co-migrating with authentic baccatin III (Rf = 0.45 for the standard) was isolated and analyzed by radio-HPLC to reveal the new radioactive product described above. Extracts of *E. coli* transformed with empty vector controls did not yield detectable product when assayed by identical methods.

KK 4/13/03

Figures 10A and 10B

~~**Figure 10:**~~ combined reverse-phase HPLC-chemical ionization MS (mass spectrometry) analysis of (spectrum A) the biosynthetic product (Rt = 8.6 ± 0.1 minutes) generated by recombinant acetyl transferase with 10-deaceylbaccatin III and acetyl CoA as co-substrates, and of (spectrum B) authentic baccatin III (Rt = 8.6 ± 0.1 minutes). The diagnostic mass spectral fragments are at m/z 605 ($M + NH_4^+$), 587 (MH^+), 572 ($MH^+ - CH_3$), 527 ($MH^+ - CH_3COOH$), and 509 ($MH^+ - (CH_3COOH + H_2O)$). For preparation of recombinant enzyme and product isolation, see **Figure 8** legend.

DETAILED DESCRIPTION

Definitions

Mammal: This term includes both humans and non-human mammals. Similarly, the term "patient" includes both humans and veterinary subjects.

Taxoid: A "taxoid" is a chemical based on the Taxane ring structure as described in Kinston et al., *Progress in the Chemistry of Organic Natural Products*, Springer-Verlag, 1993.

Isolated: An "isolated" biological component (such as a nucleic acid or protein or organelle) is a component that has been substantially separated or purified away from other biological components in the cell of the organism in which the component naturally occurs, i.e., other chromosomal and extra-chromosomal DNA, RNA, proteins, and organelles. Nucleic acids and proteins that have been "isolated" include nucleic acids and proteins purified by standard purification methods. The term also embraces nucleic acids and proteins prepared by recombinant expression in a host cell, as well as chemically synthesized nucleic acids.

Orthologs: An "ortholog" is a gene that encodes a protein that displays a function that is similar to a gene derived from a different species.

Homologs: "Homologs" are two nucleotide sequences that share a common ancestral sequence and diverged when a species carrying that ancestral sequence split into two species.

Purified: The term "purified" does not require absolute purity; rather, it is intended as a relative term. Thus, for example, a purified enzyme or nucleic acid preparation is one in which the subject protein or nucleotide, respectively, is at a higher concentration than the protein or nucleotide would be in its natural environment within an organism. For example, a preparation of an enzyme can be considered as purified if the enzyme content in the preparation represents at least 50% of the total protein content of the preparation.

Vector: A "vector" is a nucleic acid molecule as introduced into a host cell, thereby producing a transformed host cell. A vector may include nucleic acid sequences, such as an origin of replication, that permit the vector to replicate in a host cell. A vector may also include one or more screenable markers, selectable markers, or reporter genes and other genetic elements known in the art.

Transformed: A "transformed" cell is a cell into which a nucleic acid molecule has been introduced by molecular biology techniques. As used herein, the term "transformation" encompasses all techniques by which a nucleic acid molecule might be introduced into such a cell, including transfection with a viral vector, transformation with a plasmid vector, and introduction of naked DNA by electroporation, lipofection, and particle gun acceleration.

DNA construct: The term "DNA construct" is intended to indicate any nucleic acid molecule of cDNA, genomic DNA, synthetic DNA, or RNA origin. The term "construct" is intended to indicate a nucleic acid segment that may be single- or double-stranded, and that may be based on a complete or partial naturally occurring nucleotide sequence encoding one or more of the transacylase genes of the present invention. It is understood that such nucleotide sequences include intentionally manipulated nucleotide sequences, e.g., subjected to site-directed mutagenesis, and sequences that are degenerate as a result of the genetic code. All degenerate nucleotide sequences are included within the scope of the invention so long as the transacylase encoded by the nucleotide sequence maintains transacylase activity as described below.

Recombinant: A "recombinant" nucleic acid is one having a sequence that is not naturally occurring in the organism in which it is expressed, or has a sequence made by

an artificial combination of two otherwise-separated, shorter sequences. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques. "Recombinant" is also used to describe nucleic acid molecules that have been artificially manipulated, but contain the same control sequences and coding regions that are found in the organism from which the gene was isolated.

Specific binding agent: A "specific binding agent" is an agent that is capable of specifically binding to the transacylases of the present invention, and may include polyclonal antibodies, monoclonal antibodies (including humanized monoclonal antibodies) and fragments of monoclonal antibodies such as Fab, F(ab')2 and Fv fragments, as well as any other agent capable of specifically binding to the epitopes on the proteins.

cDNA (complementary DNA): A "cDNA" is a piece of DNA lacking internal, non-coding segments (introns) and regulatory sequences that determine transcription. cDNA is synthesized in the laboratory by reverse transcription from messenger RNA extracted from cells.

ORF (open reading frame): An "ORF" is a series of nucleotide triplets (codons) coding for amino acids without any termination codons. These sequences are usually translatable into respective polypeptides.

Operably linked: A first nucleic acid sequence is "operably linked" with a second nucleic acid sequence whenever the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two protein-coding regions, in the same reading frame.

Probes and primers: Nucleic acid probes and primers may be prepared readily based on the amino acid sequences and nucleic acid sequences provided by this invention. A "probe" comprises an isolated nucleic acid attached to a detectable label or reporter

molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. Methods for labeling and guidance in the choice of labels appropriate for various purposes are discussed in, e.g., Sambrook et al. (ed.), *Molecular Cloning: A Laboratory Manual* 2nd ed., vol. 1-3, cold Spring Harbor Laboratory Press, cold Spring Harbor, NY, 1989, and Ausubel et al. (ed.) *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, New York (with periodic updates), 1987.

"Primers" are short nucleic acids, preferably DNA oligonucleotides 10 nucleotides or more in length. A primer may be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, and then extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR), or other nucleic-acid amplification methods known in the art.

Methods for preparing and using probes and primers are described, for example, in references such as Sambrook et al. (ed.), *Molecular Cloning: A Laboratory Manual,* 2nd ed., vol. 1-3, cold Spring Harbor Laboratory Press, cold Spring Harbor, NY, 1989; Ausubel et al. (ed.), *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, New York (with periodic updates), 1987; and Innis et al., *PCR Protocols: A Guide to Methods and Applications*, Academic Press: San Diego, 1990. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, © 1991, Whitehead Institute for Biomedical Research, Cambridge, MA). One of skill in the art will appreciate that the specificity of a particular probe or primer increases with the length of the probe or primer. Thus, for example, a primer comprising 20 consecutive nucleotides will anneal to a target having a higher specificity than a corresponding primer of only 15 nucleotides. Thus, in order to obtain greater specificity, probes and primers may be selected that comprise, for example, 10, 20, 25, 30, 35, 40, 50 or more consecutive nucleotides.

Sequence identity: The similarity between two nucleic acid sequences or between two amino acid sequences is expressed in terms of the level of sequence identity shared between the sequences. Sequence identity is typically expressed in terms of percentage identity; the higher the percentage, the more similar the two sequences.

Methods for aligning sequences for comparison are well known in the art. Various programs and alignment algorithms are described in: Smith & Waterman, *Adv. Appl. Math.* **2:**482, 1981; Needleman & Wunsch, *J. Mol. Biol.* **48:**443, 1970; Pearson & Lipman, *Proc. Natl. Acad. Sci. USA* **85:**2444, 1988; Higgins & Sharp, *Gene* **73:**237-244, 1988; Higgins & Sharp, *CABIOS* **5:**151-153, 1989; Corpet et al., *Nucleic Acids Research* **16:**10881-10890, 1988; Huang, et al., *CABIOS* **8:**155-165, 1992; and Pearson et al., *Methods in Molecular Biology* **24:**307-331, 1994. Altschul et al., *J. Mol. Biol.* **215:**403-410, 1990, presents a detailed consideration of sequence alignment methods and homology calculations.

The National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST™, Altschul et al.. *J. Mol. Biol.* **215:**403-410, 1990) is available from several sources, including the National Center for Biotechnology Information (NCBI, Bethesda, MD) and on the Internet, for use in connection with the sequence-analysis programs blastp, blastn, blastx, tblastn and tblastx. BLAST™ can be accessed on the internet at http://www.ncbi.nlm.nih.gov/BLAST/. A description of how to determine sequence identity using this program is available on the internet at http://www.ncbi.nlm.nih.gov/BLAST/blast_help.html.

For comparisons of amino acid sequences of greater than about 30 amino acids, the "Blast 2 sequences" function of the BLAST™ program is employed using the default BLOSUM62 matrix set to default parameters, (gap existence cost of 11, and a per residue gap cost of 1). When aligning short peptides (fewer than around 30 amino acids), the alignment should be performed using the Blast 2 sequences function, employing the PAM30 matrix set to default parameters (open gap 9, extension gap 1 penalties). Proteins with even greater similarity to the reference sequences will show increasing percentage identities when assessed by this method, such as at least 45%, at least 50%, at least 60%, at least 80%, at least 85%, at least 90%, or at least 95% sequence identity.

Transacylase (an older name for acyltransferase) activity: Enzymes exhibiting transacylase activity are capable of transferring acyl groups, forming either esters or amides, by catalyzing reactions in which an acyl group that is linked to a carrier (acyl-carrier) is transferred to a reactant, thus forming an acyl group linked to the reactant (acyl-reactant).

Transacylases: Transacylases are enzymes that display transacylase activity as described *supra*. However, all transacylases do not recognize the same carriers and

reactants. Therefore, transacylase enzyme-activity assays must utilize different substrates and reactants depending on the specificity of the particular transacylase enzyme. One of ordinary skill in the art will appreciate that the assay described below is a representative example of a transacylase activity assay, and that similar assays can be used to test transacylase activity directed towards different substrates and reactants.

Substantial similarity: A first nucleic acid is "substantially similar" to a second nucleic acid if, when optimally aligned (with appropriate nucleotide deletions or gap insertions) with the other nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about, for example, 50%, 75%, 80%, 85%, 90% or 95% of the nucleotide bases. Sequence similarity can be determined by comparing the nucleotide sequences of two nucleic acids using the BLAST™ sequence analysis software (blastn) available from The National Center for Biotechnology Information. Such comparisons may be made using the software set to default settings (expect = 10, filter = default, descriptions = 500 pairwise, alignments = 500, alignment view = standard, gap existence cost = 11, per residue existence = 1, per residue gap cost = 0.85). Similarly, a first polypeptide is substantially similar to a second polypeptide if they show sequence identity of at least about 75%-90% or greater when optimally aligned and compared using BLAST software (blastp) using default settings.

II. Characterization of acetyl CoA:taxa-4(20),11(12)-dien-5α-ol *O*-acetyl transacylase

A. Enzyme Purification and Library construction

Biochemical studies have indicated that the third specific intermediate of the Taxol™ biosynthesis pathway is taxa-4(20),11(12)-dien-5∀-yl acetate, because this metabolite serves as a precursor of a series of polyhydroxy taxanes en route to the end-product (Hezari and Croteau, *Planta Medica* **63**:291-295, 1997). The responsible enzyme, taxadienol acetyl transacylase, that converts taxadienol to the C5-acetate ester is, thus, an important candidate for cDNA isolation for the purpose of overexpression in relevant producing organisms to increase Taxol™ yield (Walker et al., *Arch. Biochem. Biophys*. **364**:273-279, 1999).

This enzyme has been partially purified and characterized with respect to reaction parameters (Walker et al., *Arch. Biochem. Biophys.* **364**:273-279, 1999); however, the published fractionation protocol does not yield a pure protein suitable for amino acid microsequencing that is required for an attempt at reverse genetic cloning of the gene. [It is also important to note that the gene has no homologs or orthologs (i.e., other terpenoid or isoprenoid *O*-acetyl transacylases) in the databases to permit similarity-based cloning approaches.]

Using methyl jasmonate-induced *Taxus canadensis* cells as an enriched enzyme source, a new isolation and purification protocol (see **Figure 3**, and protocol described *infra*) was developed to efficiently yield homogeneous protein for microsequencing. Although the protein was N-blocked and failed to yield peptides that could be internally sequenced by V8 (endoproteinase Glu-C, Roche Molecular Biochemical, Nutley, New Jersey) proteolysis or cyanogen bromide (CNBr) cleavage, treatment with endolysC (endoproteinase Lys-C, Roche Molecular Biochemical, Nutley, New Jersey) and trypsin yielded a mixture of peptides. Five of these could be separated by high-performance liquid chromatography (HPLC) and verified by mass spectrometry (MS), and yielded sequence information useful for a cloning effort (**Figure 2**).

For cDNA library construction, a stable, methyl jasmonate-inducible *T. cuspidata* suspension cell line was chosen for mRNA isolation because the production of Taxol™ was highly inducible in this system (which permits the preparation of a suitable subtractive library, if necessary). The mixing of experimental protocols as used with different *Taxus* species is not a significant limitation, since all *Taxus* species are known to be very closely related and are considered by several taxonomists to represent geographic variants of the basic species *T. baccata* (Bolsinger and Jaramillo, *Silvics of Forest Trees of North America* (revised), Pacific Northwest Research Station, USDA, p. 17, Portland, OR, 1990; and Voliotis, *Isr. J. Botany.* **35**:47-52, 1986). Thus, the genes encoding geranylgeranyl diphosphate synthase and taxadiene synthase (early steps of Taxol™ biosynthesis) from *T. canadensis* and *T. cuspidata* evidence only very minor sequence differences. Hence, a method was developed for the isolation of high-quality mRNA from *Taxus* cells (Qiagen, Valencia, California) and this material was employed for cDNA library construction using a commercial kit which is available from Stratagene, La Jolla, California.

B. Reverse Genetic Cloning

Of the five tryptic peptides that were sequenced (**Figure 2**), peptide SEQ ID NOs: 30, 31, and 33 were found to exhibit some similarity to the sequences of the only two other plant acetyl transacylases that have been documented, namely, deacetylvindoline *O*-acetyl transacylase involved in indole alkaloid biosynthesis (St. Pierre et al., *Plant J.* **14**:703-713, 1998) and benzyl alcohol *O*-acetyl transacylase involved in the biosynthesis of aromatic esters of floral scent (Dudareva et al., *Plant J.* **14**:297-304, 1998). Lesser resemblance was found to a putative aromatic *O*-benzoyl transacylase of plant origin (Yang et al., *Plant Mol. Biol.* **35**:777-789, 1997). Of the five peptide sequences (**Figure 2**), SEQ ID NO: 30 was most suitable for primer design based on codon degeneracy considerations, and two such forward degenerate primers, AT-FOR1 (SEQ ID NO: 34) and AT-FOR2 (SEQ ID NO: 35), were synthesized (**Figure 4**). A search of the database with the tryptic peptide ILVYYPPFAGR (SEQ ID NO: 30) revealed two possible variants of this sequence among several gene entries of known and unknown function (these entries are listed in Table 1). consideration of these distantly related sequences allowed the design of two additional forward degenerate primers (AT-FOR3 (SEQ ID NO: 36) and AT-FOR4 (SEQ ID NO: 37)), and permitted identification of a distal consensus sequence from which a degenerate reverse primer (AT-REV1 (SEQ ID NO: 38)) was designed (**Figure 4**). (An alignment of the *Taxus* sequences with the extant database sequence entries of Table 1 illustrates the lack of significant homology between the *Taxus* sequences and any previously described genes.)

Table 1

Database (GenBank) sequences used for peptide comparisons. For alignment, see **Figure 6**; for placement in dendrogram, see **Figure 7**. The accession number is followed by a two-letter code indicating genus and species (AT, *Arabidopsis thaliana*; CM, *Cucumis melo*; CR, *Catharanthus roseus*; DC, *Dianthus caryophyllus*; CB, *Clarkia breweri*; NT, *Nicotiana tabacum*).

Accession No.	Protein Identification No.	Function
AC000103_AT	g2213627	unknown; from genomic sequence for *Arabidopsis thaliana* BAC F21J9
AC000103_AT	g2213628	unknown; from genomic sequence for *A. thaliana* BAC F21J9
AF002109_AT	g2088651	unknown; hypersensitivity-related gene 201 isolog
AC002560_AT	g2809263	unknown; from genomic sequence for *A. thaliana* BAC F21B7
AC002986_AT	g3152598	unknown; similarity to C2-HC type zinc finger protein C.e-MyT1 gb/U67079 from *C. elegans* and to hypersensitivity-related gene 201 isolog T28M21.14 from *A. thaliana* BAC
AC002392_AT	g3176709	putative anthranilate N-hydroxycinnamoyl/benzoyltransferase
AL031369_AT	g3482975	unknown; putative protein
Z84383_AT	g2239083	hydroxycinnamoyl:benzoyl-CoA:anthranilate N-hydroxycinnamoyl:benzoyl transferase
Z97338_AT	g2244896	unknown; similar to HSR201 protein *N. tabacum*
Z97338_AT	g2244897	unknown; hypothetical protein
AL049607_AT	g4584530	unknown; putative protein
AF043464_CB	g3170250	acetyl CoA:benzylalcohol acetyl transferase
Z70521_CM	g1843440	unknown; expressed during ripening of melon (*Cucumis melo* L.) fruits
AF053307_CR	g4091808	deacetylvindoline 4-*O*-acetyl transferase
AC004512_DC	g3335350	unknown; similar to gb/Z84386 anthranilate N-hydroxycinnamoyl/benzoyltransferase from *Dianthus caryophyllus*
X95343_NT	g1171577	unknown; hypersensitive reaction in tobacco

PCR amplifications were performed using each combination of forward and reverse primers, and induced *Taxus* cell library cDNA as a target. The amplifications produced, by cloning and sequencing, twelve related but distinct amplicons (each ca. 900

bp) having origins from the various primers (Table 2). These amplicons are designated "Probe 1" through "Probe 12," and their nucleotide and deduced amino acid sequences are listed as SEQ ID NOs: 1-24, respectively.

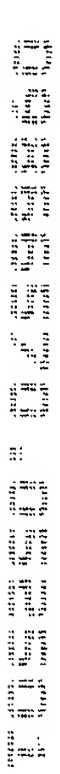

Table 2

Primer combinations, amplicons and acquired genes. The parentheses and brackets are used to designate the primer pair used and the corresponding frequency at which that primer pair amplified the probe.

Primer Pair	Amplicon			Acquired Gene	
	Size (bp)	Frequency	Designation	Designation	Function
AT-FOR1/AT-REV1 (AT-FOR2/AT-REV1)	920	7/12 (12/31)	Probe 1	TAX1 (full-length) SEQ ID NO: 27; SEQ ID NO: 28	taxadienol acetyl transferase
(Figure 4)	SEQ ID NO: 1; SEQ ID NO: 2			TAX2 (full-length) SEQ ID NO: 25; SEQ ID NO: 26	unknown
AT-FOR1/AT-REV1 (AT-FOR2/AT-Rev1)	920	7/12 (2/31)	Probe 2	Probe 2 was not used, but likely would have acquired TAX2 because the sequence corresponds directly to this gene.	-----
(Figure 4)	SEQ ID NO: 3; SEQ ID NO: 4				
AT-FOR4/AT-REV1	903	2/29	Probe 3	-----	-----
(Figure 4)	SEQ ID NO: 5; SEQ ID NO: 6				
AT-FOR3/AT-REV1	908	1/29	Probe 4	-----	-----
(Figure 4)	SEQ ID NO: 7; SEQ ID NO: 8			-----	-----
AT-FOR4/AT-REV1	908	1/32	Probe 5	TAX5 (full-length) SEQ ID NO: 49; SEQ ID NO: 50	unknown
(Figure 4)	SEQ ID NO: 9; SEQ ID NO: 10			-----	-----
AT-FOR2/AT-REV1 (AT-FOR3/AT-REV1) [AT-FOR4/AT-REV1]	911	8/32 (1/29) [1/32]	Probe 6	TAX6 (full-length) SEQ ID NO: 44; Seq. ID No: 45	10-deacetylbaccatin III-10-*O*-acetyl transferase
(Figure 4)	SEQ ID NO: 11; SEQ ID NO: 12			-----	-----
AT-FOR3/AT-REV1	968	6/29	Probe 7	TAX7 (full-length) SEQ ID NO: 51; SEQ ID NO: 52	unknown
(Figure 4)	SEQ ID NO: 13; SEQ ID NO: 14			-----	-----
AT-FOR3/AT-REV1 (AT-FOR4/AT-REV1)	908	1/29 (2/32)	Probe 8	-----	-----
(Figure 4)	SEQ ID NO: 15; SEQ ID NO: 16			-----	-----
AT-FOR2/AT-REV1 (AT-FOR3/AT-REV1)	908	1/32 (5/29)	Probe 9	-----	-----
(Figure 4)	SEQ ID NO: 17; SEQ ID NO: 18			-----	-----
AT-FOR4/AT-REV1	911	2/32	Probe 10	TAX10 (full-length) SEQ ID NO: 53; SEQ ID NO: 54	unknown
(Figure 4)	SEQ ID NO: 19; SEQ ID NO: 20			-----	-----
AT-FOR4/AT-REV1	920	1/32	Probe 11	-----	-----
(Figure 4)	SEQ ID NO: 21; SEQ ID NO: 22			-----	-----
AT-FOR3/AT-REV1 (AT-FOR4/AT-REV1)	908	3/29 (1/32)	Probe 12	TAX12 (full-length) SEQ ID NO: 55; SEQ ID NO: 56	unknown
(Figure 4)	SEQ ID NO: 23; SEQ ID NO: 24			-----	-----
	TAX13 does not appear to directly correspond to any of the above listed Probes			TAX13 (full-length) SEQ ID NO: 57; SEQ ID NO: 58	unknown

Notably, Probe 1, derived from the primers AT-FOR1 (SEQ ID NO: 34) and AT-REV1 (SEQ ID NO: 38), amplified a ~900 bp DNA fragment encoding, with near identity, the proteolytic peptides corresponding to SEQ ID NOs: 31-33 of the purified protein. These results suggested that the amplicon Probe 1 represented the target gene for taxadienol acetyl transacylase. Probe 1 was then ^{32}P-labeled and employed as a hybridization probe in a screen of the methyl jasmonate-induced *T. cuspidata* suspension cell λZAP II™ cDNA library. Standard hybridization and purification procedures ultimately led to the isolation of three full-length, unique clones designated TAX1, TAX2, and TAX6 (SEQ ID NOS: 27, 25, and 44, respectively).

C. Sequence Analysis and Functional Expression

Clone TAX1 bears an open reading frame of 1317 nucleotides (nt; SEQ ID NO: 27)) and encodes a deduced protein of 439 amino acids (aa; SEQ ID NO: 28) with a calculated molecular weight of 49,079 kDa. Clone TAX2 bears an open reading frame of 1320 nt (SEQ ID NO:25) and encodes a deduced protein of 440 aa (SEQ ID NO:26) with a calculated molecular weight of 50,089 kDa. Clone TAX6 bears an open reading frame of 1320 nt (SEQ ID NO: 44) and encodes a deduced protein of 440 aa (SEQ ID NO: 45) with a calculated molecular weight of 49,000 kDa.

The sizes of TAX1 and TAX2 are consistent with the molecular weight of the native taxadienol transacetylase (MW ~50,000) determined by gel-permeation chromatography (Walker et al., *Arch. Biochem. Biophys.* **364**:273-279, 1999) and SDS polyacrylamide gel electrophoresis (SDS-PAGE). The deduced amino acid sequences of both TAX1 and TAX2 also remotely resemble those of other acetyl transacylases (50-56% identity; 64-67% similarity) involved in different pathways of secondary metabolism in plants (St. Pierre et al., *Plant J.* **14**:703-713, 1998; and Dudareva et al., *Plant J.* **14**:297-304, 1998). When compared to the amino acid sequence information from the tryptic peptide fragments, TAX1 exhibited a very close match (91% identity), whereas TAX2 exhibited conservative differences (70% identity).

The TAX6 calculated molecular weight of 49,052 kDa is consistent with that of the native TAX6 protein (~50 kDa), determined by gel permeation chromatography, indicating the protein to be a functional monomer, and is very similar to the size of the related, monomeric taxadien-5α-ol transacetylase (MW = 49,079). The acetyl CoA:10-

deacetylbacctin III-10-*O*-acetyl transferase from *Taxus cuspidata* appears to be substantially different in size from the acetyl CoA:10-hydroxytaxane-*O*-acetyl transferase recently isolated from *Taxus chinensis* and reported at a molecular weight of 71,000 (Menhard and Zenk, *Phytochemistry* **50**:763-774, 1999).

The deduced amino acid sequence of TAX6 resembles that of TAX1 (64 % identity; 80 % similarity) and those of other acetyl transferases (56-57 % identity; 65-67 % similarity) involved in different pathways of secondary metabolism in plants (Dudareva et al., *Plant J.* **14**:297-304, 1998; St-Pierre et al., *Plant J.* **14**:703-713, 1998). Additionally, TAX6 possesses the HXXXDG (SEQ ID NO: 48) (residues H162, D166, and G167, respectively) motif found in other acyl transferases (Brown et al., *J. Biol. Chem.* **269**:19157-19162, 1994; Carbini and Hersh, *J. Neurochem.* **61**:247-253, 1993; Hendle et al., *Biochemistry* **34**:4287-4298, 1995; and Lewendon et al., *Biochemistry* **33**:1944-1950, 1994); this sequence element has been suggested to function in acyl group transfer from acyl CoA to the substrate alcohol (St. Pierre et al., *Plant J.* **14**:703-713, 1998).

To determine the identity of the putative taxadienol acetyl transacylase, TAX1, TAX2, and TAX6 were subcloned in-frame into the expression vector pCWori+ (Barnes, *Methods Enzymol.* **272**:3-14, 1996) and expressed in *E. coli* JM109 cells. The transformed bacteria were cultured and induced with isopropyl ∃-D-thiogalactoside (IPTG), and cell-free extracts were prepared and evaluated for taxadienol acetyl transacylase activity using the previously developed assay procedures (Walker et al., *Arch. Biochem. Biophys.* **364**:273-279, 1999). Clone TAX1 (corresponding directly to Probe 1) expressed high levels of taxadienol acetyl transacylase activity (20% conversion of substrate to product), as determined by radiochemical analysis; the product of this recombinant enzyme was confirmed as taxadienyl-5∀-yl acetate by gas chromatography-mass spectrometry (GC-MS) (**Figure 5**). Clone TAX2 did not express taxadienol acetyl transacylase activity and was inactive with the [^{3}H]taxadienol and acetyl CoA co-substrates. However, the clone TAX2 may encode an enzyme for a step later in the Taxol™ biosynthetic pathway (TAX2 has been shown to correspond to Probe 2). Neither of the recombinant proteins expressed from TAX1 or TAX2 was capable of acetylating the advanced Taxol™ precursor 10-deacetyl baccatin III to baccatin III. Thus, based on the demonstration of functionally expressed activity, and the resemblance of the recombinant enzyme in substrate specificity and other physical and chemical properties to

the native form, clone TAX1 was confirmed to encode the *Taxus* taxadienol acetyl transacylase.

Additionally, the heterologously expressed TAX6 was partially purified by anion-exchange chromatography (*O*-diethylaminoethylcellulose, Whatman, Clifton, NJ) and ultrafiltration (Amicon Diaflo YM 10 membrane, Millipore, Bedford, MA) to remove interfering hydrolases from the bacterial extract, and the recombinant enzyme was determined to catalyze the conversion of 10-deacetylbaccatin III to baccatin III; the latter is the last diterpene intermediate in the Taxol™ (paclitaxel) biosynthetic pathway. The optimum pH for TAX6 was determined to be 7.5, with half-maximal velocities at pH 6.4 and 7.8. The K_m values for 10-deacetylbaccatin III and acetyl CoA were determined to be 10 μM and 8 μM, respectively, by Lineweaver-Burk analysis (for both plots $R^2 = 0.97$). These kinetic constants for TAX6 are comparable to the taxa-4(20),11(12)-dien-5α-ol acetyl transferase possessing K_m values for taxadienol and acetyl CoA of 4 μM and 6 μM, respectively. The TAX6 appears to acetylate the 10-hydroxyl group of taxoids with a high degree of regioselectivity, since the enzyme does not acetylate the 1β-, 7β-, or 13α-hydroxyl groups of 10-deacetylbaccatin III, nor does it acetylate the 5α-hydroxyl group of taxa-4(20),11(12)-dien-5α-ol.

III. Other Transacylases of the Taxol™ Pathway

The protocol described above yielded twelve related amplicons. Initial use of the first and second amplicons as probes for screening the cDNA library allowed for the isolation and characterization of taxadienol 5-*O*-acetyl transacylase. In addition to this first confirmed taxadienol 5-*O*-acetyl transacylase (TAX1), there are at least four additional transacylation steps in the Taxol™ biosynthetic pathway represented by the 2-debenzoyl baccatin III-2-*O*-benzoyl transacylase, the 10-deacetylbaccatin III-10-*O*-acetyl transacylase, the baccatin III-13-*O*-phenylisoserinyl transacylase, and the debenzoyltaxol-N-benzoyl transacylase. The close relationship between the nucleic acid sequences of the twelve amplicons indicates that the remaining amplicon sequences represent partial nucleic acid sequences of the other transacylases in the Taxol™ pathway. Hence, the above-described protocol enables full-length versions of these Taxol™ transacylases to be obtained. The following discussion relating to Taxol™ transacylases refers to taxadienol 5-*O*-acetyl transacylase, as well as the remaining transacylases of the Taxol™ pathway.

Furthermore, one of skill in the art will appreciate that the remaining transacylases can be tested easily for enzymatic activity using functional assays with the appropriate taxoid substrates, see for example the assay for taxoid C10 transacylase described in Menhard and Zenk, *Phytochemistry* **50**:763-774, 1999.

IV. Isolating a Gene Encoding acetyl CoA:taxa-4(20),11(12)-dien-5α-ol *O*-acetyl transacylase

A. Experimental Overview

A newly designed isolation and purification method is described below for the preparation of homogeneous taxadien-5∀-ol acetyl transacylase from *Taxus canadensis*. The purified protein was N-terminally blocked, thereby requiring internal amino acid microsequencing of fragments generated by proteolytic digestion. Peptide fragments so generated were purified by HPLC and sequenced, and one suitable sequence was used to design a set of degenerate PCR primers. Several primer combinations were employed to amplify a series of twelve related, gene-specific DNA sequences (Probes 1-12). Nine of these gene-specific sequences were used as hybridization probes to screen an induced *Taxus cuspidata* cell cDNA library. This strategy allowed for the successful isolation of eight full-length transacylase cDNA clones. The identity of one of these clones was confirmed by sequence matching to the peptide fragments described above and by heterologous functional expression of transacylase activity in *Escherichia coli*.

B. Culture of Cells

Initiation, propagation and induction of *Taxus sp.* cell cultures, reagents, procedures for the synthesis of substrates and standards, and general methods for transacylase isolation, characterization and assay have been previously described (Hefner et al., *Arch. Biochem. Biophys.* **360**:62-75, 1998; and Walker et al., *Arch. Biochem. Biophys.* **364**:273-279, 1999). Since all designated *Taxus* species are considered to be closely related subspecies (Bolsinger and Jaramillo, *Silvics of Forest Trees of North America* (revised), Pacific Northwest Research Station, USDA, Portland, OR, 1990; and Voliotis, *Isr. J. Botany* **35**:47-52, 1986), the *Taxus* cell sources were chosen for operational considerations because only minor sequence differences and/or allelic variants between proteins and genes of the various "species" were expected. Thus, *Taxus canadensis* cells were chosen as the source of transacetylase because they express

transacetylase at high levels, and *Taxus cuspidata* cells were selected for cDNA library construction because they produce Taxol™ at high levels.

C. Isolation and Purification of the Enzyme

No related terpenol transacylase genes are available in the databases (see below) to permit homology-based cloning. Hence, a protein-based (reverse genetic) approach to cloning the target transacetylase was required. This reverse genetic approach required obtaining a partial amino acid sequence, generating degenerate primers, amplifying a portion of cDNA using PCR, and using the amplified fragment as a probe to detect the correct clone in a cDNA library.

Unfortunately, the previously described partial protein purification protocol, including an affinity chromatography step, did not yield pure protein for amino acid microsequencing, nor did the protocol yield protein in useful amounts, or provide a sufficiently simplified SDS-PAGE banding pattern to allow assignment of the transacetylase activity to a specific protein (Walker et al., *Arch. Biochem. Biophys.* **364:**273-279, 1999). Furthermore, numerous variations on the affinity chromatography step, as well as the earlier anion exchange and hydrophobic interaction chromatography steps, failed to improve the specific activity of the preparations due to the instability of the enzyme upon manipulation. Also, a five-fold increase in the scale of the preparation resulted in only marginally improved recovery (generally <5% total yield accompanied by removal of >99% of total starting protein). Furthermore, because the enzyme could not be purified to homogeneity, and attempts to improve stability by the addition of polyols (sucrose, glycerol), reducing agents ($Na_2S_2O_5$, ascorbate, dithiothreitol, Ǝ-mercaptoethanol), and other proteins (albumin, casein) were also not productive (Walker et al., *Arch. Biochem. Biophys.* **364**:273-279, 1999), this approach had to be abandoned.

To overcome the problem described above, the following isolation and purification procedure was used. The purity of the taxadienol acetyl transacylase after each fractionation step was assessed by SDS-PAGE according to Laemmli (Laemmli, *Nature* **227**:680-685, 1970); quantification of total protein after each purification step was carried out by the method of Bradford, *Analytical Biochem.* **72:**248-254, 1976, or by Coommassie Blue staining, and transacylase activity was assessed using the methods described in Walker et al., *Arch. Biochem. Biophys.* **364**:273-279, 1999.

Procedures for protein staining have been described (Wray et al., *Anal. Biochem.* **118**:197-203, 1991). The preparation of the *T. canadensis* cell-free extracts and all subsequent procedures were performed at 0-4°C unless otherwise noted. Cells (40 g batches) were frozen in liquid nitrogen and thoroughly pulverized for 1.5 minutes using a mortar and pestle. The resulting frozen powder was transferred to 225 mL of ice cold 30 mM HEPES buffer (pH 7.4) containing 3 mM dithiothreitol (DTT), XAD-4 polystyrene resin (12 g) and polyvinylpolypyrrolidone (PVPP, 12 g) to adsorb low molecular weight resinous and phenolic compounds. The slurry was slowly stirred for 30 minutes, and the mixture was filtered through four layers of cheese cloth to remove solid absorbents and particulates. The filtrate was centrifuged at 7000 *g* for 30 minutes to remove cellular debris, then at 100,000 *g* for 3 hours, followed by 0.2-µm filtration to yield a soluble protein fraction (in ~200 mL buffer) used as the enzyme source.

Jul C16 The soluble enzyme fraction was subjected to ultrafiltration (DIAFLO™ YM 30 membrane, Millipore, Bedford, Massachusetts) to concentrate the fraction from 200 mL to 40 mL and to selectively remove proteins of molecular weight lower than the taxadien-5∀-ol acetyl transacylase (previously established at 50,000 Da in Walker et al., *Arch. Biochem. Biophys.* **364**:273-279, 1999). Using a peristaltic pump, the concentrate (40 mL) was applied (2 mL/minute) to a column of *O*-diethylaminoethylcellulose (2.8 X 10 cm, Whatman DE-52, Fairfield, New Jersey) that had been equilibrated with "equilibration buffer" (30 mM HEPES buffer (pH 7.4) containing 3 mM DTT). After washing with 60 mL of equilibration buffer to remove unbound material, the proteins were eluted with a step gradient of the same buffer containing 50 mM (25 mL), 125 mM (50 mL), and 200 mM (50 mL) NaCl.

Jul C17 The fractions were assayed as described previously (Walker et al., *Arch. Biochem. Biophys.* **364**:273-279, 1999), and those containing taxadien-5∀-ol acetyl transacylase activity (125-mM and 200-mM fractions) were combined (100 mL, ~160 mM) and diluted to 5 mM NaCl (160 mL) by ultrafiltration (DIAFLO™ YM 30 membrane, Millipore, Bedford, Massachusetts) and repeated dilution with 30 mM HEPES buffer (pH 7.4) containing 3 mM DTT.

Further purification was effected by high-resolution anion-exchange and hydroxyapatite chromatography run on a Pharmacia FPLC system coupled to a 280-nm effluent detector. The preparation described above was applied to a preparative anion-exchange column (10 X 100 mm, Source 15Q, Pharmacia Biotech., Piscataway, New

Jersey) that was previously washed with "wash buffer" (30 mM HEPES buffer (pH 7.4) containing 3 mM DTT) and 1 M NaCl, and then equilibrated with wash buffer (without NaCl). After removing unbound material, the applied protein was eluted with a linear gradient of 0 to 200 mM NaCl in equilibration buffer (215 mL total volume; 3 mL/minute) (see **Figure 3A**). Fractions containing transacetylase activity (eluting at ~80 mM NaCl) were combined and diluted to 5 mM NaCl by ultrafiltration using 30 mM HEPES buffer (pH 7.4) containing 3 mM DTT as diluent, as described above. The desalted protein sample (70 mL) was loaded onto an analytical anion-exchange column (5 X 50 mm, Source 15Q, Pharmacia Biotech., Piscataway, New Jersey) that was washed and equilibrated as before. The column was developed using a shallow, linear salt gradient with elution to 200 mM NaCl (275 mL total volume, 1.5 mL/minute, 3.0 mL fractions). The taxadienol acetyl transacylase eluted at ~55-60 mM NaCl (see **Figure 3B**), and the appropriate fractions were combined (15 mL), reconstituted to 45 mL in 30 mM HEPES buffer (pH 6.9) and applied to a ceramic hydroxyapatite column (10 X 100 mm, Bio-Rad Laboratories, Hercules, California) that was previously washed with 200 mM sodium phosphate buffer (pH 6.9) and then equilibrated with an "equilibration buffer" (30 mM HEPES buffer (pH 6.9) containing 3 mM DTT (without sodium phosphate)). The equilibration buffer was used to desorb weakly associated material, and the bound protein was eluted by a gradient from 0 to 40 mM sodium phosphate in equilibration buffer (125 mL total volume, at 3.0 mL/minute, 3.0 mL fractions) (see **Figure 3C**). The fractions containing the highest activity, eluting over 27 mL at 10 mM sodium phosphate, were combined and shown by SDS-PAGE to yield a protein of ~95% purity (a minor contaminant was present at ~35 kDa, see **Figure 3D**). The level of transacylase activity was measured after each step in the isolation and purification protocol described above. The level of activity recovered is shown in Table 3.

Table 3

Summary of taxadien-5α-ol *O*-acetyl transferase purification from *Taxus* cells.

	Total activity (pkat)	Total Protein (mg)	Specific Activity (pkat/mg protein)	Purification (fold)
Crude extract	302	1230	0.25	1
YM30 ultrafiltration	136	98	1.4	5.6
DE-52	122	69	1.8	7.2
YM30 ultrafiltration	54	55	1.0	4
Source 15Q (10 X 100 mm)	47	3	16	63

	Total activity (pkat)	Total Protein (mg)	Specific Activity (pkat/mg protein)	Purification (fold)
YM30 ultrafiltration	19	2.6	7.3	29
Source 15Q (5 X 50 mm)	13	0.12	108	400
Hydroxyapatite	10	0.05	200	800

D. Amino Acid Microsequencing of Taxadienol Acetyl Transacylase

The purified protein from multiple preparations as described above (>95% pure, ~100 pmol, 50 µg) was subjected to preparative SDS-PAGE (Laemmli, *Nature* **227**:680-685, 1970). The protein band at 50 kDa, corresponding to the taxadienol acetyl transacylase, was excised. Whereas treatment with V8 protease or treatment with cyanogen bromide (CNBr) failed to yield sequencable peptides, *in situ* proteolysis with endolysC (Caltech Sequence/Structure Analysis Facility, Pasadena, CA) and trypsin (Fernandez et al., *Anal. Biochem.* **218**:112-118, 1994) yielded a number of peptides, as determined by HPLC, and several of these were separated, verified by mass spectrometry (Fernandez et al., *Electrophoresis* **19**:1036-1045, 1998), and subjected to Edman degradative sequencing, from which five distinct and unique amino acid sequences (designated SEQ ID NOs: 29-33) were obtained (**Figure 2**).

E. cDNA Library construction and Related Manipulations

A cDNA library was constructed from mRNA isolated from *T. cuspidata* suspension culture cells that had been induced to maximal Taxol™ production with methyl jasmonate for 16 hours. An optimized protocol for the isolation of total RNA from *T. cuspidata* cells was developed empirically using a buffer containing 100 mM Tri-HCl (pH 7.5), 4 M guanidine thiocyanate, 25 mM EDTA and 14 mM ∃-mercaptoethanol. Cells (1.5 g) were disrupted at 0-4°C using a Polytron™ ultrasonicator (Kinematica AG, Switzerland; 4 X 15 second bursts at power setting 7), the resulting homogenate was adjusted to 2% (v/v) Triton X-100 and allowed to stand 15 minutes on ice. An equal volume of 3 M sodium acetate (pH 6.0) was then added, and the mixed solution was incubated on ice for an additional 15 minutes, followed by centrifugation at 15,000 *g* for 30 minutes at 4°C. The resulting supernatant was mixed with 0.8 volume of isopropanol and allowed to stand on ice for 5 minutes, followed by centrifugation at 15,000 *g* for 30

minutes at 4°C. The resulting pellet was dissolved in 8 mL of 20 mM Tris-HCl (pH 8.0) containing 1 mM EDTA, adjusted to pH 7.0 by addition of 2 mL of 2 M NaCl in 250 mM MOPS buffer (pH 7.0), and total RNA was recovered by passing this solution over a nucleic acid isolation column (Qiagen, Valencia, California) following the manufacturer's instructions. Poly(A)+ mRNA was then purified from total RNA by chromatography on oligo(dT) beads (Oligotex™ mRNA Kit, Qiagen), and this material was used to construct a library using the λZAPII™ cDNA synthesis kit and Gigapack™ III gold packaging kit from Stratagene, La Jolla, California, by following the manufacturer's instructions.

Unless otherwise stated, standard methods were used for DNA manipulations and cloning (Sambrook et al. (ed.), *Molecular Cloning: A Laboratory Manual* 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and for PCR amplification procedures (Innis et al., *PCR Protocols: A Guide to Methods and Applications*, Academic Press, New York, 1990). DNA was sequenced using Amplitaq™ (Hoffmann-La Roche INC., Nutley, New Jersey) DNA polymerase and cycle sequencing (fluorescence sequencing) on an ABI Prism™ 373 DNA Sequencer. The *E. coli* strains XL1-Blue and XL1-Blue MRF' (Stratagene, La Jolla, California) were used for routine cloning of PCR products and for cDNA library construction, respectively. E. coli XL1-Blue MRF'cells were used for *in vivo* excision of purified pBluescript SK from positive plaques and the excised plasmids were used to transform *E. coli* SOLR cells.

F. Degenerate Primer Design and PCR Amplification

Due to codon degeneracy, only one sequence of the five tryptic peptide fragments obtained (SEQ ID NO: 30 of **Figure 2**) was suitable for PCR primer construction. Two such degenerate forward primers, designated AT-FOR1 (SEQ ID NO: 34) and AT-FOR2 (SEQ ID NO: 35), were designed based on this sequence (**Figure 4**). Using the NCBI Blast 2.0 database searching program (Genetics computer Group, Program Manual for the Wisconsin Package, version 9, Genetics computer Group, 575 Science Drive, Madison, WI, 1994) to search for this sequence element among the few defined transacylases of plant origin (St. Pierre et al., *Plant J.* **14**:703-713, 1998; Dudareva et al., *Plant J.* **14**:297-304, 1998; and Yang et al., *Plant Mol. Bio.* **35:**777-789, 1997), and the many deposited sequences of unknown function, allowed the identification of two possible sequence variants of this element (FYPFAGR (SEQ ID NO: 39) and YYPLAGR (SEQ ID NO: 40))

from which two additional degenerate forward primers, designated AT-FOR3 (SEQ ID NO: 36) and AT-FOR4 (SEQ ID NO: 37), were designed (**Figure 4**). The sequences employed for this comparison are listed in Table 1. Using this range of functionally defined and undefined sequences, conserved regions were sought for the purpose of designing a degenerate reverse primer (the distinct lack of similarity of the *Taxus* sequences to genes in the database can be appreciated by reference to **Figure 6**), from which one such consensus sequence element (DFGWGKP) (SEQ ID NO: 41) was noted, and was employed for the design of the reverse primer AT-REV1 (SEQ ID NO: 38) (**Figure 4**). This set of four forward primers and one reverse primer incorporated a varied number of inosines, and ranged from 72- to 216-fold degeneracy. The remaining four proteolytic peptide fragment sequences (SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33 of **Figure 2**) were not only less suitable for primer design, but they were not found (by NCBI BLAST™ searching) to be similar to other related sequences, thus suggesting that these represented more specific sequence elements of the *Taxus* transacetylase gene.

Each forward primer (150 μM) and the reverse primer (150 μM) were used in separate PCR reactions performed with *Taq* polymerase (3 U/100 μL reaction containing 2 mM $MgCl_2$) and employing the induced *T. cuspidata* cell cDNA library (10^8 PFU) as template under the following conditions: 94°C for 5 minutes, 32 cycles at 94°C for 1 minute, 40°C for 1 minute and 74°C for 2 minutes and, finally, 74°C for 5 minutes. The resulting amplicons (regions amplified by the various primer combinations) were analyzed by agarose gel electrophoresis (Sambrook et al. (ed.), *Molecular Cloning: A Laboratory Manual* 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989) and the products were extracted from the gel, ligated into pCR TOPOT7 (Invitrogen, Carlsbad, California), and transformed into *E. coli* TOPIOF' cells (Invitrogen, Carlsbad, California). Plasmid DNA was prepared from individual transformants and the inserts were fully sequenced.

The combination of primers AT-FOR1 (SEQ ID NO: 34) and AT-REV1 (SEQ ID NO: 38) yielded a 900-bp amplicon. Cloning and sequencing of the amplicon revealed two unique sequences designated "Probe 1" (SEQ ID NO: 1) and "Probe 2" (SEQ ID NO: 3) (Table 2). The results with the remaining primer combinations are provided in Table 2.

G. Library screening

Four separate library-screening experiments were designed using various combinations of the radio-labeled amplicons (Probes 1-12, described *supra*) as probes. Use of radio-labeled Probe 1 (SEQ ID NO: 1), led to the identification of TAX1 (SEQ ID NO: 27) and TAX2 (SEQ ID NO: 25), and use of radio-labeled Probe 6 (SEQ ID NO: 11) led to the identification of TAX6 (SEQ ID NO: 44). A probe consisting of a mixture of radio-labeled Probe 10 (SEQ ID NO: 19) and Probe 12 (SEQ ID NO: 23) led to the identification of TAX10 (SEQ ID NO: 44) and TAX12 (SEQ ID NO: 55). Finally, a probe containing a mixture of radio-labeled Probes 3, 4, 5, 7, and 9 led to the identification of TAX5, TAX 7, and TAX13 (SEQ ID NOs. 49, 51, and 57, respectively). Details of these individual library-screening experiments are provided below.

The identification of TAX1 (SEQ ID NO: 27) and TAX2 (SEQ ID NO: 25) was accomplished using 1 μg of Probe 1 (SEQ ID NO: 1) that had been amplified by PCR, the resulting amplicon was gel-purified, randomly labeled with [∀-^{32}P]CTP (Feinberg and Vogelstein, *Anal. Biochem.* **137**:216-217, 1984), and used as a hybridization probe to screen membrane lifts of 5 X 10^5 plaques grown in *E. coli* XL1-Blue MRF'. Phage DNA was cross-linked to the nylon membranes by autoclaving on fast cycle 3-4 minutes at 120°C. After cooling, the membranes were washed 5 minutes in 2 X SSC, then 5 minutes in 6 X SSC (containing 0.5% SDS, 5 X Denhardt's reagent, 0.5 g Ficoll (Type 400, Pharmacia, Piscataway, New Jersey), 0.5 g polyvinylpyrrolidone (PVP-10), and 0.5 g bovine serum albumin (Fraction V, Sigma, Saint Louis, Missouri) in 100 mL total volume). Hybridization was then performed for 20 hours at 68°C in 6 X SSC, 0.5% SDS and 5 X Denhardt's reagent. The nylon membranes were then washed two times for 5 minutes in 2 X SSC with 0.1% SDS at 25°C, and then washed 2 X 30 minutes with 1 X SSC and 0.1% SDS at 68°C. After washing, the membranes were exposed for 17 hours to Kodak (Rochester, New York) XAR film at -70°C (Sambrook et al. (ed.), *Molecular Cloning: A Laboratory Manual*, 2nd ed., vol. 1-3, cold Spring Harbor Laboratory Press, cold Spring Harbor, NY, 1989).

Of the plaques exhibiting positive signals (~600 total), 60 were purified through two additional rounds of hybridization. Purified λZAPII clones were excised *in vivo* as pBluescript II SK(-) phagemids and transformed into *E. coli* SOLR cells (Stratagene, La Jolla, California). The size of each cDNA insert was determined by PCR using T3 and T7 promoter primers, and size-selected inserts (>1.5 kb) were partially sequenced from

both ends to sort into unique sequence types and to acquire full-length versions of each (by further screening with a newly designed 5'-probe, if necessary).

The same basic screening protocol, as illustrated by the results provided below, can be repeated with all of the probes described in Table 2, with the goal of acquiring the full range of full-length, in-frame putative transacylase clones for test of function by expression in *E. coli*. In the case of Probe 1 (SEQ ID NO: 1), two unique full-length clones, designated TAX1 (SEQ ID NO: 27 and SEQ ID NO: 28) and TAX2 (SEQ ID NO: 25 and SEQ ID NO: 26), were isolated.

An additional transacylase, TAX6 (SEQ ID NO: 44), was identified by using 40 ng of radio-labeled Probe 6 (SEQ ID NO: 11) to screen the *T. cuspidata* library. This full-length clone was 99% identical to Probe 6 (SEQ ID NO: 11) and 99% identical to the deduced amino acid sequence of Probe 6 (SEQ ID NO: 12), indicating that the probe had located its cognate.

Using 40 ng of radio-labeled Probe 10 (SEQ ID NO: 19) and 40 ng of radio-labeled Probe 12 (SEQ ID NO: 23) led to the identification of the full-length transacylases TAX10 (SEQ ID NO: 53 and SEQ ID NO:54) and TAX12 (SEQ ID NO:55 and SEQ ID NO: 56) in separate hybridization screening experiments.

Use of a probe mixture containing about 6 ng each of Probes 3, 4, 5, 7, and 8 (SEQ ID NOs. 5, 7, 9, 13, and 15, respectively) randomly labeled with [α-^{32}P]CTP (Feinberg and Vogelstein, *Anal. Biochem.* **137:**216-2 17, 1984) resulted in the identification of full-length transacylases TAX5 (SEQ ID NO: 49) and TAX7 (SEQ ID NO: 51), which correspond to Probes 5 (SEQ ID NO: 9) and 7 (SEQ ID NO: 13), respectively. An additional full-length transacylase, TAX13 (SEQ ID NO: 57) was also identified, however, this transacylase does not correspond to any of the Probes identified in Table 2.

H. cDNA Expression in E. coli

Full-length insert fragments of the relevant plasmids are excised and subcloned in-frame into the expression vector pCWori+ (Barnes, *Methods Enzymol.* **272:**3-14, 1996). This procedure may involve the elimination of internal restriction sites and the addition of appropriate 5'- and 3'-restriction sites for directional ligation into the expression vector using standard PCR protocols (Innis et al., *PCR Protocols: A Guide to Methods and Applications*, Academic Press: San Diego, 1990) or commercial kits such as the Quick

Change Mutagenesis System (Stratagene, La Jolla, California). For example, the full-length transacylase corresponding to probe 6 (SEQ ID NO: 11) was obtained using the primer set (5'-GGGAATTCCATATGGCAGGCTCAACAGAATTTGTGG-3' (SEQ ID NO: 46) and 3'-GTTTATACATTGATTCGGAACTAGATCTGATC-5' (SEQ ID NO:47)) to amplify the putative full-length acetyl transferase gene and incorporate *NdeI* and *XbaI* restriction sites at the 5'- and 3'-termini, respectively, for directional ligation into vector pCWori+ (Barnes, *Methods Enzymol.* **272:**3-14, 1996). All recombinant pCWori+ plasmids are confirmed by sequencing to insure that no errors have been introduced by the polymerase reactions, and are then transformed into *E. coli* JM109 by standard methods.

Isolated transformants for each full-length insert are grown to A_{600} = 0.5 at 37°C in 50 mL Luria-Bertani medium supplemented with 50 μg ampicillin/mL, and a 1-mL inoculum added to a large scale (100 mL) culture of Terrific Broth (6 g bacto-tryptone, DIFCO Laboratories, Spark, Maryland, 12 g yeast extract, EM Science, Cherryhill, New Jersey, and 2 mL glycerol in 500 mL water) containing 50 μg ampicillin/mL and thiamine HCl (320 mM) and grown at 28°C for 24 hours. Approximately 24 hours after induction with 1 mM isopropyl ∃-D-thiogalactoside (IPTG), the bacterial cells are harvested by centrifugation, disrupted by sonication in assay buffer consisting of 30 mM potassium phosphate (pH 7.4), or 25 mM MOPSO (pH 7.4), followed by centrifugation to yield a soluble enzyme preparation that can be assayed for transacylase activity.

I. Enzyme assay

A specific assay for acetyl CoA:taxa-4(20),11(12)-dien-5∀-ol *O*-acetyl transacylase has been described previously (Walker et al., *Arch. Biochem. Biophys.* **364**:273-279, 1999, herein incorporated by reference). Generally the assay for taxoid acyltransacylases involves the CoA-dependent acyl transfer from acetyl CoA (or other acyl or aroyl CoA ester) to a taxane alcohol, and the isolation and chromatographic separation of the product ester for confirmation of structure by GC-MS (or HPLC-MS) analysis. For another example of such an assay, see Menhard and Zenk, *Phytochemistry* **50**:763-774, 1999.

The activity of TAX6 (SEQ ID NO: 45) was assayed under standard conditions described in Walker et al., *Arch. Biochem. Biophys.* **364**:273-279, 1999, with 10-deacetylbaccatin III (400 μM, Hauser Chemical Research Inc., Boulder, CO) and [2-

^{3}H]acetyl CoA (0.45 μCi, 400 μM (NEN, Boston, MA)) as co-substrates. The TAX6 (SEQ ID NO: 45) enzyme preparation yielded a single product from reversed-phase radio-HPLC analysis, with a retention time of 7.0 minutes (coincident radio and UV traces) corresponding exactly to that of authentic baccatin III (generously provided by Dr. David Bailey of Hauser Chemical Research Inc., Boulder, CO) (**Figure 9**). The identity of the biosynthetic product was further verified as baccatin III by combined LC-MS (liquid chromatography-mass spectrometry) analysis (**Figure 10**), which demonstrated the identical retention time (8.6 x 0.1 minute) and mass spectrum for the product and authentic standard. Finally, a sample of the biosynthetic product, purified by silica gel analytical TLC, gave a ^{1}H-NMR spectrum identical to that of authentic baccatin III, confirming the enzyme as 10-deacetylbaccatin III-10-*O*-acetyl transferase (TAX6 (SEQ ID NO: 45)) and also confirming that the corresponding gene had been isolated.

EXAMPLES

1. Transacylase Protein and Nucleic acid Sequences

As described above, the invention provides transacylases and transacylase-specific nucleic acid sequences. With the provision herein of these transacylase sequences, the polymerase chain reaction (PCR) may now be utilized as a preferred method for identifying and producing nucleic acid sequences encoding the transacylases. For example, PCR amplification of the transacylase sequences may be accomplished either by direct PCR from a plant cDNA library or by Reverse-Transcription PCR (RT-PCR) using RNA extracted from plant cells as a template. Transacylase sequences may be amplified from plant genomic libraries, or plant genomic DNA. Methods and conditions for both direct PCR and RT-PCR are known in the art and are described in Innis et al., *PCR Protocols: A Guide to Methods and Applications*, Academic Press: San Diego, 1990.

The selection of PCR primers is made according to the portions of the cDNA (or gene) that are to be amplified. Primers may be chosen to amplify small segments of the cDNA, the open reading frame, the entire cDNA molecule or the entire gene sequence. Variations in amplification conditions may be required to accommodate primers of differing lengths; such considerations are well known in the art and are discussed in Innis et al., *PCR Protocols: A Guide to Methods and Applications*, Academic Press: San Diego, 1990; Sambrook et al. (ed.), *Molecular Cloning: A Laboratory Manual* 2nd ed.,

vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel et al. (ed.) *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, New York (with periodic updates), 1987. By way of example, the cDNA molecules corresponding to additional transacylases may be amplified using primers directed towards regions of homology between the 5' and 3' ends of the TAX1 and TAX2 sequences. Example primers for such a reaction are:

primer 1: 5' CCT CAT CTT TCC CCC ATT GAT AAT 3' (SEQ ID NO: 42)

primer 2: 5' AAA AAG AAA ATA ATT TTG CCA TGC AAG 3' (SEQ ID NO: 43)

These primers are illustrative only; it will be appreciated by one skilled in the art that many different primers may be derived from the provided nucleic acid sequences. Re-sequencing of PCR products obtained by these amplification procedures is recommended to facilitate confirmation of the amplified sequence and to provide information on natural variation between transacylase sequences. Oligonucleotides derived from the transacylase sequence may be used in such sequencing methods.

Oligonucleotides that are derived from the transacylase sequences are encompassed within the scope of the present invention. Preferably, such oligonucleotide primers comprise a sequence of at least 10-20 consecutive nucleotides of the transacylase sequences. To enhance amplification specificity, oligonucleotide primers comprising at least 15, 20, 25, 30, 35, 40, 45 or 50 consecutive nucleotides of these sequences may also be used.

A. Transacylases in Other Plant Species

Orthologs of the transacylase genes are present in a number of other members of the *Taxus* genus. With the provision herein of the transacylase nucleic acid sequences, the cloning by standard methods of cDNAs and genes that encode transacylase orthologs in these other species is now enabled. As described above, orthologs of the disclosed transacylase genes have transacylase biological activity and are typically characterized by possession of at least 50% sequence identity counted over the full length alignment with the amino acid sequence of the disclosed transacylase sequences using the NCBI Blast 2.0 (gapped blastp set to default parameters). Proteins with even greater similarity to the

reference sequences will show increasing percentage identities when assessed by this method, such as at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 90%, or at least 95% sequence identity.

Both conventional hybridization and PCR amplification procedures may be utilized to clone sequences encoding transacylase orthologs. Common to both of these techniques is the hybridization of probes or primers that are derived from the transacylase nucleic acid sequences. Furthermore, the hybridization may occur in the context of Northern blots, Southern blots, or PCR.

Direct PCR amplification may be performed on cDNA or genomic libraries prepared from any of various plant species, or RT-PCR may be performed using mRNA extracted from plant cells using standard methods. PCR primers will comprise at least 10 consecutive nucleotides of the transacylase sequences. One of skill in the art will appreciate that sequence differences between the transacylase nucleic acid sequence and the target nucleic acid to be amplified may result in lower amplification efficiencies. To compensate for this, longer PCR primers or lower annealing temperatures may be used during the amplification cycle. Where lower annealing temperatures are used, sequential rounds of amplification using nested primer pairs may be necessary to enhance specificity.

For conventional hybridization techniques the hybridization probe is preferably conjugated with a detectable label such as a radioactive label, and the probe is preferably at least 10 nucleotides in length. As is well known in the art, increasing the length of hybridization probes tends to give enhanced specificity. The labeled probe derived from the transacylase nucleic acid sequence may be hybridized to a plant cDNA or genomic library and the hybridization signal detected using methods known in the art. The hybridizing colony or plaque (depending on the type of library used) is then purified and the cloned sequence contained in that colony or plaque is isolated and characterized.

Orthologs of the transacylases alternatively may be obtained by immunoscreening of an expression library. With the provision herein of the disclosed transacylase nucleic acid sequences, the enzymes may be expressed and purified in a heterologous expression system (e.g., *E. coli*) and used to raise antibodies (monoclonal or polyclonal) specific for transacylases. Antibodies may also be raised against synthetic peptides derived from the transacylase amino acid sequence presented herein. Methods of raising antibodies are well known in the art and are described generally in Harlow and Lane, *Antibodies, A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring, N.Y. 1988. Such antibodies

can then be used to screen an expression cDNA library produced from a plant. This screening will identify the transacylase ortholog. The selected cDNAs can be confirmed by sequencing and enzyme activity assays.

B. Taxol™ Transacylase Variants

With the provision of the transacylase amino acid sequences (SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 45, 50, 52, 54, 56, and 58) and the corresponding cDNA (SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 44, 49, 51, 53, 55, and 57), variants of these sequences can now be created.

Variant transacylases include proteins that differ in amino acid sequence from the transacylase sequences disclosed, but that retain transacylase biological activity. Such proteins may be produced by manipulating the nucleotide sequence encoding the transacylase using standard procedures such as site-directed mutagenesis or the polymerase chain reaction. The simplest modifications involve the substitution of one or more amino acids for amino acids having similar biochemical properties. These so-called "conservative substitutions" are likely to have minimal impact on the activity of the resultant protein. Table 4 shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative substitutions.

Table 4

Original Residue	conservative Substitutions
ala	ser
arg	lys
asn	gln; his
asp	glu
cys	ser
gln	asn
glu	asp
gly	pro
his	asn; gln
ile	leu; val
leu	ile; val
lys	arg; gln; glu
met	leu; ile
phe	met; leu; tyr
ser	thr
thr	ser
trp	tyr
tyr	trp; phe
val	ile; leu

More substantial changes in enzymatic function or other features may be obtained by selecting substitutions that are less conservative than those in Table 4, i.e., selecting residues that differ more significantly in their effect on maintaining: (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation; (b) the charge or hydrophobicity of the molecule at the target site; or (c) the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in protein properties will be those in which: (a) a hydrophilic residue, e.g., seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g., leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine. The effects of these amino acid substitutions or deletions or additions may be assessed for transacylase derivatives by analyzing the ability of the derivative proteins to catalyse the conversion of one Taxol™ precursor to another Taxol™precursor.

Variant transacylase cDNA or genes may be produced by standard DNA mutagenesis techniques, for example, M13 primer mutagenesis. Details of these techniques are provided in Sambrook et al. (ed.), *Molecular Cloning: A Laboratory Manual* 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989, Ch. 15. By the use of such techniques, variants may be created that differ in minor ways from the transacylase cDNA or gene sequences, yet that still encode a protein having transacylase biological activity. DNA molecules and nucleotide sequences that are derivatives of those specifically disclosed herein and that differ from those disclosed by the deletion, addition, or substitution of nucleotides while still encoding a protein having transacylase biological activity are comprehended by this invention. In their simplest form, such variants may differ from the disclosed sequences by alteration of the coding region to fit the codon usage bias of the particular organism into which the molecule is to be introduced.

Alternatively, the coding region may be altered by taking advantage of the degeneracy of the genetic code to alter the coding sequence in such a way that, while the nucleotide sequence is substantially altered, it nevertheless encodes a protein having an amino acid sequence identical or substantially similar to the disclosed transacylase amino acid sequences. For example, the fifteenth amino acid residue of the TAX2 (SEQ ID NO: 26) is alanine. This is encoded in the open reading frame (ORF) by the nucleotide codon triplet GCG. Because of the degeneracy of the genetic code, three other nucleotide codon triplets -- GCA, GCC, and GCT -- also code for alanine. Thus, the nucleotide sequence of the ORF can be changed at this position to any of these three codons without affecting the amino acid composition of the encoded protein or the characteristics of the protein. Based upon the degeneracy of the genetic code, variant DNA molecules may be derived from the cDNA and gene sequences disclosed herein using standard DNA mutagenesis techniques as described above, or by synthesis of DNA sequences. Thus, this invention also encompasses nucleic acid sequences that encode the transacylase protein but that vary from the disclosed nucleic acid sequences by virtue of the degeneracy of the genetic code.

Variants of the transacylase may also be defined in terms of their sequence identity with the transacylase amino acid and nucleic acid sequences described *supra*. As described above, transacylases have transacylase biological activity and share at least 60% sequence identity with the disclosed transacylase sequences. Nucleic acid sequences that encode such proteins may readily be determined simply by applying the genetic code

to the amino acid sequence of the transacylase, and such nucleic acid molecules may be readily produced by assembling oligonucleotides corresponding to portions of the sequence.

As previously mentioned, another method of identifying variants of the transacylases is nucleic acid hybridization. Nucleic acid molecules that are derived from the transacylase cDNA and gene sequences include molecules that hybridize under various conditions to the disclosed Taxol™ transacylase nucleic acid molecules, or fragments thereof. Generally, hybridization conditions are classified into categories, for example very high stringency, high stringency, and low stringency. The conditions for probes that are about 600 base pairs or more in length are provided below in three corresponding categories.

<u>Very High Stringency (detects sequences that share 90% sequence identity)</u>						
Hybridization	in	5x	SSC	at	65°C	16 hours
Wash twice	in	2x	SSC	at	room temp.	15 minutes each
Wash twice	in	0.5x	SSC	at	65°C	20 minutes each
<u>High Stringency (detects sequences that share 80% sequence identity or greater)</u>						
Hybridization	in	5x	SSC	at	65°C	16 hours
Wash twice	in	2x	SSC	at	room temp.	20 minutes each
Wash once	in	1x	SSC	at	55°C	30 minutes each
<u>Low Stringency (detects sequences that share greater than 50% sequence identity)</u>						
Hybridization	in	6x	SSC	at	room temp.	16 hours
Wash twice	in	3x	SSC	at	room temp. (20-21°C)	20 minutes each

The sequences encoding the transacylases identified through hybridization may be incorporated into transformation vectors and introduced into host cells to produce transacylase.

2. Introduction of Transacylases into Plants

After a cDNA (or gene) encoding a protein involved in the determination of a particular plant characteristic has been isolated, standard techniques may be used to

express the cDNA in transgenic plants in order to modify the particular plant characteristic. The basic approach is to clone the cDNA into a transformation vector, such that the cDNA is operably linked to control sequences (e.g., a promoter) directing expression of the cDNA in plant cells. The transformation vector is then introduced into plant cells by any of various techniques (e.g., electroporation) and progeny plants containing the introduced cDNA are selected. Preferably all or part of the transformation vector stably integrates into the genome of the plant cell. That part of the transformation vector that integrates into the plant cell and that contains the introduced cDNA and associated sequences for controlling expression (the introduced "transgene") may be referred to as the recombinant expression cassette.

Selection of progeny plants containing the introduced transgene may be made based upon the detection of an altered phenotype. Such a phenotype may result directly from the cDNA cloned into the transformation vector or may be manifested as enhanced resistance to a chemical agent (such as an antibiotic) as a result of the inclusion of a dominant selectable marker gene incorporated into the transformation vector.

Successful examples of the modification of plant characteristics by transformation with cloned cDNA sequences are replete in the technical and scientific literature. Selected examples, which serve to illustrate the knowledge in this field of technology include:

U.S. Patent No. 5,571,706 ("Plant Virus Resistance Gene and Methods")

U.S. Patent No. 5,677,175 ("Plant Pathogen Induced Proteins")

U.S. Patent No. 5,510,471 ("Chimeric Gene for the Transformation of Plants")

U.S. Patent No. 5,750,386 ("Pathogen-Resistant Transgenic Plants")

U.S. Patent No. 5,597,945 ("Plants Genetically Enhanced for Disease Resistance")

U.S. Patent No. 5,589,615 ("Process for the Production of Transgenic Plants with Increased Nutritional Value Via the Expression of Modified 2S Storage Albumins")

U.S. Patent No. 5,750,871 ("Transformation and Foreign Gene Expression in Brassica Species")

U.S. Patent No. 5,268,526 ("Overexpression of Phytochrome in Transgenic Plants")

U.S. Patent No. 5,262,316 ("Genetically Transformed Pepper Plants and Methods for their Production")

U.S. Patent No. 5,569,831 ("Transgenic Tomato Plants with Altered Polygalacturonase Isoforms")

These examples include descriptions of transformation vector selection, transformation techniques, and the construction of constructs designed to over-express the introduced cDNA. In light of the foregoing and the provision herein of the transacylase amino acid sequences and nucleic acid sequences, it is thus apparent that one of skill in the art will be able to introduce the cDNAs, or homologous or derivative forms of these molecules, into plants in order to produce plants having enhanced transacylase activity. Furthermore, the expression of one or more transacylases in plants may give rise to plants having increased production of Taxol™ and related compounds.

A. Vector construction, Choice of Promoters

A number of recombinant vectors suitable for stable transfection of plant cells or for the establishment of transgenic plants have been described including those described in Weissbach and Weissbach, *Methods for Plant Molecular Biology*, Academic Press, 1989; and Gelvin et al., *Plant and Molecular Biology Manual*, Kluwer Academic Publishers, 1990. Typically, plant-transformation vectors include one or more cloned plant genes (or cDNAs) under the transcriptional control of 5'- and 3'-regulatory sequences and a dominant selectable marker. Such plant transformation vectors typically also contain a promoter regulatory region (e.g., a regulatory region controlling inducible or constitutive, environmentally or developmentally regulated, or cell- or tissue-specific expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

Examples of constitutive plant promoters that may be useful for expressing the cDNA include: the cauliflower mosaic virus (CaMV) 35S promoter, which confers constitutive, high-level expression in most plant tissues (see, e.g., Odel et al., *Nature* **313**:810, 1985; Dekeyser et al., *Plant Cell* **2**:591, 1990; Terada and Shimamoto, *Mol. Gen. Genet.* **220**:389, 1990; and Benfey and Chua, *Science* **250**:959-966, 1990); the nopaline synthase promoter (An et al., *Plant Physiol.* **88**:547, 1988); and the octopine synthase promoter (Fromm et al., *Plant Cell* **1**:977, 1989). *Agrobacterium*-mediated transformation of *Taxus* species has been accomplished, and the resulting callus cultures have been shown to produce Taxol™ (Han et al., *Plant Science* **95**: 187-196, 1994). Therefore, it is likely that incorporation of one or more of the described transacylases

under the influence of a strong promoter (like CaMV promoter) would increase production yields of Taxol™ and related taxoids in such transformed cells.

A variety of plant-gene promoters that are regulated in response to environmental, hormonal, chemical, and/or developmental signals also can be used for expression of the cDNA in plant cells, including promoters regulated by: (a) heat (Callis et al., *Plant Physiol.* **88**:965, 1988; Ainley, et al., *Plant Mol. Biol.* **22**:13-23, 1993; and Gilmartin et al., *The Plant Cell* **4**:839-949, 1992); (b) light (e.g., the pea rbcS-3A promoter, Kuhlemeier et al., *Plant Cell* **1**:471, 1989, and the maize rbcS promoter, Schaffner and Sheen, *Plant Cell* **3**:997, 1991); (c) hormones, such as abscisic acid (Marcotte et al., *Plant Cell* **1**:969, 1989); (d) wounding (e.g., wunI, Siebertz et al., *Plant Cell* **1**:961, 1989); and (e) chemicals such as methyl jasmonate or salicylic acid (Gatz et al., *Ann. Rev. Plant Physiol. Plant Mol. Biol.* **48**:9-108, 1997).

Alternatively, tissue-specific (root, leaf, flower, and seed, for example) promoters (Carpenter et al., *The Plant Cell* **4**:557-571, 1992; Denis et al., *Plant Physiol.* **101**:1295-1304, 1993; Opperman et al., *Science* **263**:221-223, 1993; Stockhause et al., *The Plant Cell* **9**:479-489, 1997; Roshal et al., *Embo. J.* **6**:1155, 1987; Schernthaner et al., *Embo J.* **7**:1249, 1988; and Bustos et al., *Plant Cell* **1**:839, 1989) can be fused to the coding sequence to obtain a particular expression in respective organs.

Alternatively, the native transacylase gene promoters may be utilized. With the provision herein of the transacylase nucleic acid sequences, one of skill in the art will appreciate that standard molecular biology techniques can be used to determine the corresponding promoter sequences. One of skill in the art will also appreciate that less than the entire promoter sequence may be used in order to obtain effective promoter activity. The determination of whether a particular region of this sequence confers effective promoter activity may readily be ascertained by operably linking the selected sequence region to a transacylase cDNA (in conjunction with suitable 3' regulatory region, such as the NOS 3' regulatory region as discussed below) and determining whether the transacylase is expressed.

Plant-transformation vectors may also include RNA-processing signals, for example, introns, that may be positioned upstream or downstream of the ORF sequence in the transgene. In addition, the expression vectors may also include additional regulatory sequences from the 3'-untranslated region of plant genes, e.g., a 3'-terminator region to increase mRNA stability of the mRNA, such as the PI-II terminator region of potato or

the octopine or nopaline synthase (NOS) 3'-terminator regions. The native transacylase gene 3'-regulatory sequence may also be employed.

Finally, as noted above, plant-transformation vectors may also include dominant selectable marker genes to allow for the ready selection of transformants. Such genes include those encoding antibiotic-resistance genes (e.g., resistance to hygromycin, kanamycin, bleomycin, G418, streptomycin or spectinomycin) and herbicide-resistance genes (e.g., phosphinothricin acetyltransacylase).

B. Arrangement of Taxol™ transacylase Sequence in a Vector

The particular arrangement of the transacylase sequence in the transformation vector is selected according to the type of expression of the sequence that is desired.

In most instances, enhanced transacylase activity is desired, and the transacylase ORF is operably linked to a constitutive high-level promoter such as the CaMV 35S promoter. As noted above, enhanced transacylase activity may also be achieved by introducing into a plant a transformation vector containing a variant form of the transacylase cDNA or gene, for example a form that varies from the exact nucleotide sequence of the transacylase ORF, but that encodes a protein retaining transacylase biological activity.

C. Transformation and Regeneration Techniques

Transformation and regeneration of both monocotyledonous and dicotyledonous plant cells are now routine, and the appropriate transformation technique can be determined by the practitioner. The choice of method varies with the type of plant to be transformed; those skilled in the art will recognize the suitability of particular methods for given plant types. Suitable methods may include, but are not limited to: electroporation of plant protoplasts; liposome-mediated transformation; polyethylene glycol (PEG) mediated transformation; transformation using viruses; micro-injection of plant cells; micro-projectile bombardment of plant cells; vacuum infiltration; and *Agrobacterium tumefaciens* (AT) mediated transformation. Typical procedures for transforming and regenerating plants are described in the patent documents listed at the beginning of this section.

D. Selection of Transformed Plants

Following transformation and regeneration of plants with the transformation vector, transformed plants can be selected using a dominant selectable marker incorporated into the transformation vector. Typically, such a marker confers antibiotic resistance on the seedlings of transformed plants, and selection of transformants can be accomplished by exposing the seedlings to appropriate concentrations of the antibiotic.

After transformed plants are selected and grown to maturity, they can be assayed using the methods described herein to assess production levels of Taxol™ and related compounds.

3. Production of Recombinant Taxol™ transacylase in Heterologous Expression Systems

Various yeast strains and yeast-derived vectors are commonly used for the expression of heterologous proteins. For instance, *Pichia pastoris* expression systems, obtained from Invitrogen (Carlsbad, California), may be used to practice the present invention. Such systems include suitable *Pichia pastoris* strains, vectors, reagents, transformants, sequencing primers, and media. Available strains include KM71H (a prototrophic strain), SMD1168H (a prototrophic strain), and SMD1168 (a pep4 mutant strain) (Invitrogen Product Catalogue, 1998, Invitrogen, Carlsbad CA).

Non-yeast eukaryotic vectors may be used with equal facility for expression of proteins encoded by modified nucleotides according to the invention. Mammalian vector/host cell systems containing genetic and cellular control elements capable of carrying out transcription, translation, and post-translational modification are well known in the art. Examples of such systems are the well-known baculovirus system, the ecdysone-inducible expression system that uses regulatory elements from *Drosophila melanogaster* to allow control of gene expression, and the sindbis viral-expression system that allows high-level expression in a variety of mammalian cell lines, all of which are available from Invitrogen, Carlsbad, California.

The cloned expression vector encoding one or more transacylases may be transformed into any of various cell types for expression of the cloned nucleotide. Many different types of cells may be used to express modified nucleic acid molecules. Examples include cells of yeasts, fungi, insects, mammals, and plants, including transformed and non-transformed cells. For instance, common mammalian cells that

could be used include HeLa cells, SW-527 cells (ATCC deposit #7940), WISH cells (ATCC deposit #CCL-25), Daudi cells (ATCC deposit #CCL-213), Mandin-Darby bovine kidney cells (ATCC deposit #CCL-22) and Chinese hamster ovary (CHO) cells (ATCC deposit #CRL-2092). common yeast cells include *Pichia pastoris* (ATCC deposit #201178) and *Saccharomyces cerevisiae* (ATCC deposit #46024). Insect cells include cells from *Drosophila melanogaster* (ATCC deposit #CRL-10191), the cotton bollworm (ATCC deposit #CRL-9281), and *Trichoplusia ni* egg cell homoflagellates. Fish cells that may be used include those from rainbow trout (ATCC deposit #CLL-55), salmon (ATCC deposit #CRL-1681), and zebrafish (ATCC deposit #CRL-2147). Amphibian cells that may be used include those of the bullfrog, *Rana castebelana* (ATCC deposit #CLL-41). Reptile cells that may be used include those from Russell's viper (ATCC deposit #CCL-140). Plant cells that could be used include *Chlamydomonas* cells (ATCC deposit #30485), *Arabidopsis* cells (ATCC deposit #54069) and tomato plant cells (ATCC deposit #54003). Many of these cell types are commonly used and are available from the ATCC as well as from commercial suppliers such as Pharmacia (Uppsala, Sweden), and Invitrogen.

Expressed protein may be accumulated within a cell or may be secreted from the cell. Such expressed protein may then be collected and purified. This protein may then be characterized for activity and stability and may be used to practice any of the various methods according to the invention.

4. Creation of Transacylase-Specific Binding Agents

Antibodies to the transacylase enzymes, and fragments thereof, of the present invention may be useful for purification of the enzymes. The provision of the transacylase sequences allows for the production of specific antibody-based binding agents to these enzymes.

Monoclonal or polyclonal antibodies may be produced to the transacylases, portions of the transacylases, or variants thereof. Optimally, antibodies raised against epitopes on these antigens will specifically detect the enzyme. That is, antibodies raised against the transacylases would recognize and bind the transacylases, and would not substantially recognize or bind to other proteins. The determination that an antibody specifically binds to an antigen is made by any one of a number of standard immunoassay methods; for instance, Western blotting , Sambrook et al. (ed.), *Molecular Cloning: A*

Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

To determine that a given antibody preparation (such as a preparation produced in a mouse against TAX1) specifically detects the transacylase by Western blotting, total cellular protein is extracted from cells and electrophoresed on an SDS-polyacrylamide gel. The proteins are then transferred to a membrane (for example, nitrocellulose) by Western blotting, and the antibody preparation is incubated with the membrane. After washing the membrane to remove non-specifically bound antibodies, the presence of specifically bound antibodies is detected by the use of an anti-mouse antibody conjugated to an enzyme such as alkaline phosphatase; application of 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium results in the production of a densely blue-colored compound by immuno-localized alkaline phosphatase.

Antibodies that specifically detect a transacylase will, by this technique, be shown to bind substantially only the transacylase band (having a position on the gel determined by the molecular weight of the transacylase). Non-specific binding of the antibody to other proteins may occur and may be detectable as a weaker signal on the Western blot (which can be quantified by automated radiography). The non-specific nature of this binding will be recognized by one skilled in the art by the weak signal obtained on the Western blot relative to the strong primary signal arising from the specific anti-transacylase binding.

Antibodies that specifically bind to transacylases belong to a class of molecules that are referred to herein as "specific binding agents." Specific binding agents that are capable of specifically binding to the transacylase of the present invention may include polyclonal antibodies, monoclonal antibodies and fragments of monoclonal antibodies such as Fab, $F(ab')_2$ and Fv fragments, as well as any other agent capable of specifically binding to one or more epitopes on the proteins.

Substantially pure transacylase suitable for use as an immunogen can be isolated from transfected cells, transformed cells, or from wild-type cells. Concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms per milliliter. Alternatively, peptide fragments of a transacylase may be utilized as immunogens. Such fragments may be chemically synthesized using standard methods, or may be obtained by cleavage of the whole transacylase enzyme followed by purification of the desired peptide fragments. Peptides as short as three or four amino acids in length are immunogenic when presented

to an immune system in the context of a Major Histocompatibility Complex (MHC) molecule, such as MHC class I or MHC class II. Accordingly, peptides comprising at least 3 and preferably at least 4, 5, 6 or more consecutive amino acids of the disclosed transacylase amino acid sequences may be employed as immunogens for producing antibodies.

Because naturally occurring epitopes on proteins frequently comprise amino acid residues that are not adjacently arranged in the peptide when the peptide sequence is viewed as a linear molecule, it may be advantageous to utilize longer peptide fragments from the transacylase amino acid sequences for producing antibodies. Thus, for example, peptides that comprise at least 10, 15, 20, 25, or 30 consecutive amino acid residues of the amino acid sequence may be employed. Monoclonal or polyclonal antibodies to the intact transacylase, or peptide fragments thereof may be prepared as described below.

A. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to any of various epitopes of the transacylase enzymes that are identified and isolated as described herein can be prepared from murine hybridomas according to the classic method of Kohler & Milstein, *Nature* **256:**495, 1975, or a derivative method thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody-producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, *Enzymol.* **70:**419, 1980, or a derivative method thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Harlow & Lane, *Antibodies, A Laboratory Manual,* cold Spring Harbor Laboratory, New York, 1988.

B. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein,

which can be unmodified or modified, to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than other molecules and may require the use of carriers and an adjuvant. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low-titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appear to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis et al., *J. Clin. Endocrinol. Metab.* **33**:988-991, 1971.

Booster injections can be given at regular intervals, and antiserum harvested when the antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony et al., *Handbook of Experimental Immunology,* Wier, D. (ed.), Chapter 19, Blackwell, 1973. A plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/mL of serum (about 12 μM). Affinity of the antisera for the antigen is determined by preparing competitive binding curves using conventional methods.

C. Antibodies Raised by Injection of cDNA

Antibodies may be raised against the transacylases of the present invention by subcutaneous injection of a DNA vector that expresses the enzymes in laboratory animals, such as mice. Delivery of the recombinant vector into the animals may be achieved using a hand-held form of the Biolistic system (Sanford et al., *Particulate Sci. Technol.* **5**:27-37, 1987, as described by Tang et al., *Nature* (London) **356**:153-154, 1992). Expression vectors suitable for this purpose may include those that express the cDNA of the enzyme under the transcriptional control of either the human β-actin promoter or the cytomegalovirus (CMV) promoter. Methods of administering naked DNA to animals in a manner resulting in expression of the DNA in the body of the animal are well known and are described, for example, in U.S. Patent Nos. 5,620,896 ("DNA Vaccines Against Rotavirus Infections"); 5,643,578 ("Immunization by Inoculation of DNA Transcription Unit"); and 5,593,972 ("Genetic Immunization"), and references cited therein.

D. Antibody Fragments

Antibody fragments may be used in place of whole antibodies and may be readily expressed in prokaryotic host cells. Methods of making and using immunologically effective portions of monoclonal antibodies, also referred to as "antibody fragments," are well known and include those described in Better & Horowitz, *Methods Enzymol.* **178**:476-496, 1989; Glockshuber et al. *Biochemistry* **29**:1362-1367, 1990; and U.S. Patent Nos. 5,648,237 ("Expression of Functional Antibody Fragments"); No. 4,946,778 ("Single Polypeptide Chain Binding Molecules"); and No. 5,455,030 ("Immunotherapy Using Single Chain Polypeptide Binding Molecules"), and references cited therein.

5. Taxol™ Production *in vivo*

The creation of recombinant vectors and transgenic organisms expressing the vectors are important for controlling the production of transacylases. These vectors can be used to decrease transacylase production, or to increase transacylase production. A decrease in transacylase production will likely result from the inclusion of an antisense sequence or a catalytic nucleic acid sequence that targets the transacylase encoding nucleic acid sequence. Conversely, increased production of transacylase can be achieved by including at least one additional transacylase encoding sequence in the vector. These vectors can then be introduced into a host cell, thereby altering transacylase production. In the case of increased production, the resulting transacylase may be used in *in vitro* systems, as well as *in vivo* for increased production of Taxol™, other taxoids, intermediates of the Taxol™ biosynthetic pathway, and other products.

Increased production of Taxol™ and related taxoids *in vivo* can be accomplished by transforming a host cell, such as one derived from the *Taxus* genus, with a vector containing one or more nucleic acid sequences encoding one or more transacylases. Furthermore, the heterologous or homologous transacylase sequences can be placed under the control of a constitutive promoter, or an inducible promoter. This will lead to the increased production of transacylase, thus eliminating any rate-limiting effect on Taxol™ production caused by the expression and/or activity level of the transacylase.

6. Taxol™ Production *in vitro*

Currently, Taxol™ is produced by a semisynthetic method described in Hezari and Croteau, *Planta Medica* **63:**291-295, 1997. This method involves extracting 10-deacetyl-baccatin III, or baccatin III, intermediates in the Taxol™ biosynthetic pathway, and then finishing the production of Taxol™ using *in vitro* techniques. As more enzymes are identified in the Taxol™ biosynthetic pathway, it may become possible to completely synthesize Taxol™ *in vitro*, or at least increase the number of steps that can be performed *in vitro*. Hence, the transacylases of the present invention may be used to facilitate the production of Taxol™ and related taxoids in synthetic or semi-synthetic methods. Accordingly, the present invention enables the production of transgenic organisms that not only produce increased levels of Taxol™, but also transgenic organisms that produce increased levels of important intermediates, such as 10-deacetyl-baccatin III and baccatin III.

Having illustrated and described the principles of the invention in multiple embodiments and examples, it should be apparent to those skilled in the art that the invention can be modified in arrangement and detail without departing from such principles. We claim all modifications coming within the spirit and scope of the following claims.